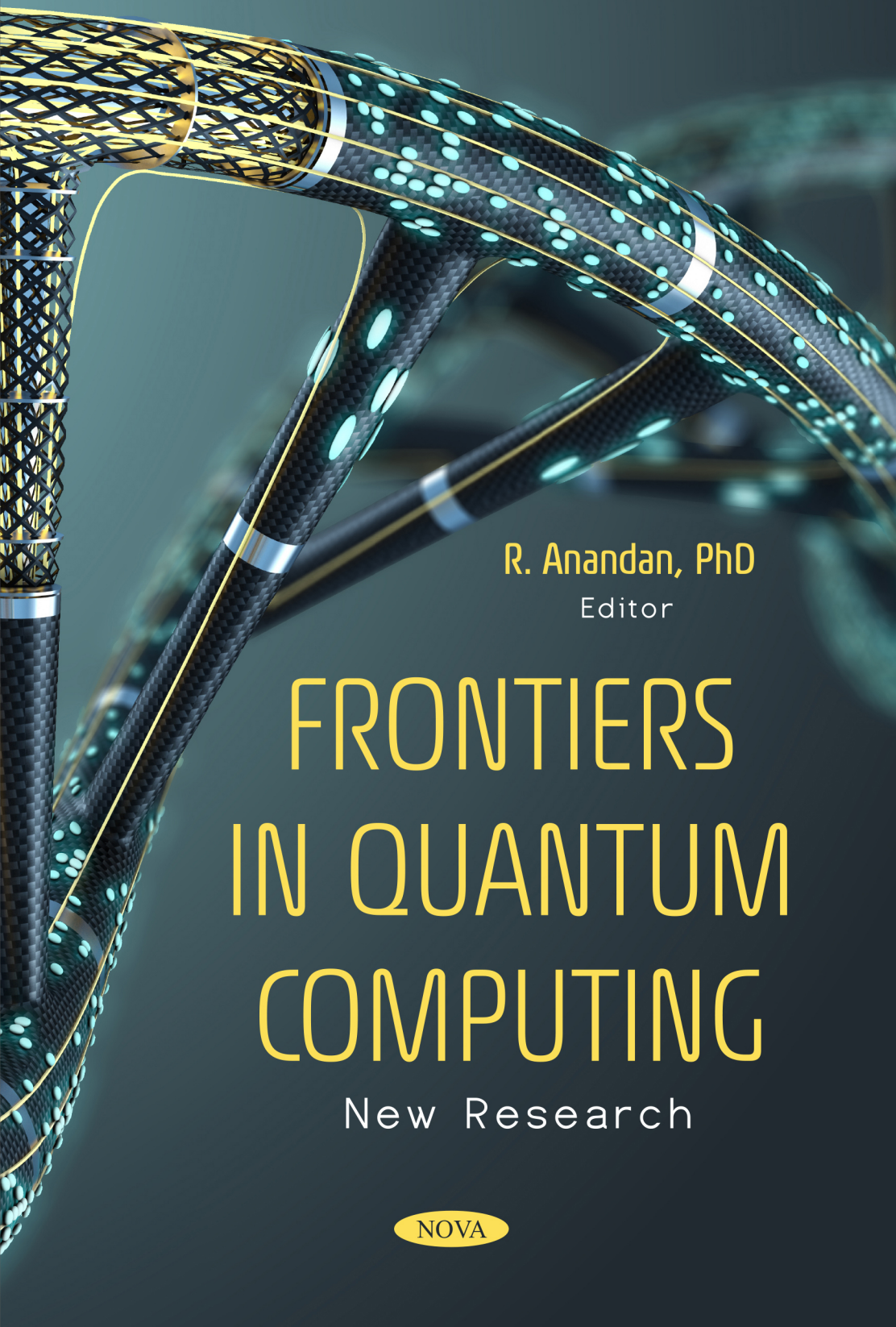
COMPUTER SCIENCE, TECHNOLOGY AND APPLICATIONS



R. Anandan, PhD
Editor

FRONTIERS IN QUANTUM COMPUTING

New Research

NOVA

Computer Science, Technology and Applications

www.novapublishers.com

No part of this digital document may be reproduced, stored in a retrieval system or transmitted in any form or by any means. The publisher has taken reasonable care in the preparation of this digital document, but makes no expressed or implied warranty of any kind and assumes no responsibility for any errors or omissions. No liability is assumed for incidental or consequential damages in connection with or arising out of information contained herein. This digital document is sold with the clear understanding that the publisher is not engaged in rendering legal, medical or any other professional services.

Computer Science, Technology and Applications

Common Visual Art in a Social Digital Age
Shalin Hai-Jew, PhD
2022. ISBN: 978-1-68507-664-1 (Hardcover)
2022. ISBN: 978-1-68507-748-8 (eBook)

Green Computing and Its Applications
Sanjay Kumar, PhD
Rohit Raja, PhD
Alok Kumar Singh Kushwaha, PhD
Raj Kumar Patra, PhD
Saurabh Kumar, PhD
(Editors)
2021. ISBN: 978-1-68507-357-2 (Hardcover)
2021. ISBN: 978-1-68507-363-3 (eBook)

Artificial Intelligence: Work, Machines and Human Interaction
Tonya Randolph
2021. ISBN: 978-1-68507-313-8 (Hardcover)
2021. ISBN: 978-1-68507-353-4 (eBook)

Advanced Decision Sciences Based on Deep Learning and Ensemble Learning Algorithms: A Practical Approach Using Python
S. Sumathi, PhD
Suresh Rajappa, PhD
L. Ashok Kumar, PhD
Surekha Paneerselvam, PhD
2021. ISBN: 978-1-68507-061-8 (Hardcover)
2021. ISBN: 978-1-68507-207-0 (eBook)

More information about this series can be found at
https://novapublishers.com/product-category/series/computer-science-technology-and-applications/

R. Anandan
Editor

Frontiers in Quantum Computing

New Research

nova
science publishers
www.novapublishers.com

Copyright © 2022 by Nova Science Publishers, Inc.
DOI: https://doi.org/10.52305/EAUU1757

All rights reserved. No part of this book may be reproduced, stored in a retrieval system or transmitted in any form or by any means: electronic, electrostatic, magnetic, tape, mechanical photocopying, recording or otherwise without the written permission of the Publisher.

We have partnered with Copyright Clearance Center to make it easy for you to obtain permissions to reuse content from this publication. Simply navigate to this publication's page on Nova's website and locate the "Get Permission" button below the title description. This button is linked directly to the title's permission page on copyright.com. Alternatively, you can visit copyright.com and search by title, ISBN, or ISSN.

For further questions about using the service on copyright.com, please contact:
Copyright Clearance Center
Phone: +1-(978) 750-8400 Fax: +1-(978) 750-4470 E-mail: info@copyright.com.

NOTICE TO THE READER

The Publisher has taken reasonable care in the preparation of this book, but makes no expressed or implied warranty of any kind and assumes no responsibility for any errors or omissions. No liability is assumed for incidental or consequential damages in connection with or arising out of information contained in this book. The Publisher shall not be liable for any special, consequential, or exemplary damages resulting, in whole or in part, from the readers' use of, or reliance upon, this material. Any parts of this book based on government reports are so indicated and copyright is claimed for those parts to the extent applicable to compilations of such works.

Independent verification should be sought for any data, advice or recommendations contained in this book. In addition, no responsibility is assumed by the Publisher for any injury and/or damage to persons or property arising from any methods, products, instructions, ideas or otherwise contained in this publication.

This publication is designed to provide accurate and authoritative information with regard to the subject matter covered herein. It is sold with the clear understanding that the Publisher is not engaged in rendering legal or any other professional services. If legal or any other expert assistance is required, the services of a competent person should be sought. FROM A DECLARATION OF PARTICIPANTS JOINTLY ADOPTED BY A COMMITTEE OF THE AMERICAN BAR ASSOCIATION AND A COMMITTEE OF PUBLISHERS.

Additional color graphics may be available in the e-book version of this book.

Library of Congress Cataloging-in-Publication Data

Names: Anandan, R., editor.
Title: Frontiers in quantum computing : new research / R. Anandan, PhD (editor), professor, head of the Department of Computer Science and Engineering, Vels Institute of Science Technology and Advanced Studies, Tamil Nadu, India.
Other titles: Frontiers in quantum computing (Nova Science Publishers : 2022)
Description: New York : Nova Science Publishers, 2022. | Series: Computer science, technology and applications | Includes bibliographical references and index. |
Identifiers: LCCN 2022018727 (print) | LCCN 2022018728 (ebook) | ISBN 9781685078164 (hardcover) | ISBN 9781685079000 (adobe pdf)
Subjects: LCSH: Quantum computing.
Classification: LCC QA76.889 .F762 2022 (print) | LCC QA76.889 (ebook) | DDC 006.3/843--dc23/eng/20220419
LC record available at https://lccn.loc.gov/2022018727
LC ebook record available at https://lccn.loc.gov/2022018728

Published by Nova Science Publishers, Inc. † New York

Contents

Preface vii

Acknowledgments xi

Chapter 1 **Programming a Quantum Computer Using Python** 1
R. Anandan, Krishnaprasanth B. V. and Sai Sucharitha K.

Chapter 2 **Blockchain-Based Quantum Key Distribution Approach** 19
M. Karthik and R. Balakrishna

Chapter 3 **Quantum Computing of PMS Using Machine Learning Algorithms for Revenue Management in Front Office Operations** 41
H. M. Moyeenudin and R. Anandan

Chapter 4 **Bernstein Vazirani and Deutsch Algorithm: Made Easy in Qiskit** 63
R. Anitha, B. Vijayalaksmi, S. Anusooya, V. Jean Shilpa and C. Rishub

Chapter 5 **Quantum Aided Deep Learning Framework for Motif Structure Prediction** 83
S. Nithinsha and D. Vinod

Chapter 6 **Quantum Behaved Translation Invariant Feature Extraction for Chromosome Classification** 115
S. Nithinsha and D. Vinod

Chapter 7 **Quantum Based Dynamic Clustering of Pharmacovigilance Data** 141
S. Nithinsha and D. Vinod

Chapter 8 **Quantum-Based Deep Learning for Multi-Level Grading of Mangoes** 171
Nirmala Gururaj, Viji Vinod and Latha Parthiban

Chapter 9 **Efficient Quantum-Based Secure Route Creation and Data Transfer in Mobile Ad-Hoc Networks Using Multi-User Co-Operative Motion Mechanism** 207
K. Pazhanisamy and Latha Parthiban

Chapter 10 **Pattern Recognition Accuracy of Echocardiogram Images Using Deep Learning Techniques** 233
Ilavazhagi Bala and S. Latha Parthiban

About the Editor 259

Index 261

Preface

It is our immense pleasure to put forth this book, *Frontiers in Quantum Computing: New Research.* The objective of this book is to communicate advancements of knowledge and help disseminate results concerning recent applications and case studies in the area of quantum computing among working professionals and professionals in education and research, covering a broad cross-section of technical disciplines.

The purpose this book is to allow students to explore knowledge in quantum computing to produce serviceable and innocuous systems as well as purposeful systems with cutting-edge technology. To yield computer systems with decent usability, developers must attempt to understand the factors that determine how people use technology.

This book will be of significance and will cater to an extensive cross-sectional and multi-disciplinary readership ranging from academics, business delegates, CEOs, communication designers, computer scientists, digital customers, e-decision makers, eLearning environment designers, industrial leaders, industry consultants, key workers, law enforcement agencies, managers, practitioners, professionals, professors, profit/non-profit e-organizations, programmers, R&D/professional research communities, security architects, stakeholders, students, support staff and university researchers/scholars of various communities, such as artificial intelligence, cyber-physical systems, ethics, robotics, safety engineering, safety-critical systems, standardization and certification digital forensics and application domain communities such as aerospace, agriculture, automotive, critical infrastructures, healthcare, manufacturing, retail, smart transports, smart cities and smart healthcare, using real case studies and projecting outcomes, showing the intricate details of quantum computing in these real-life scenarios. As a final point, this book will provide up-to-date, premeditated, and creative information to those engrossed in the field of quantum computing.

This book is divided in 10 chapters.

Chapter 1 - demonstrates how to write computer programs using the Python programming language for quantum computers.

Chapter 2 -discusses secure quantum authentication schemes. A public key mechanism and a homogeneous decryption key are employed, and the chapter explains how Morse code is utilised for the encryption process in this proposal. Morse code is a method of encrypting bits as visible light, which is necessary for the dissemination of quantum keys in blockchain technology.

Chapter 3 - describes how to evaluate the use of quantum computing in the field of front office operations of the hotel industry for better revenue management and knowing the compatibility with property management systems of luxury hotels.

Chapter 4 - highlights quantum computing concepts and implementation of basic quantum algorithms like Bernstein vazirani and Deustch algorithm in the IBM Quantum experience Qiskit tool which is an open source tool that uses Python to interface with quantum computers. Quantum computers developed by IBM can handle complicated issues that cannot be solved by efficient supercomputers.

Chapter 5 - showcases leveraging sequence, structure, and functions of proteins to identify motifs, usage motif embedding for perfect vector formation, and deep learning models to determine the motif patterns. This research work proposes a deep learning framework for DNA binding motif identification using motif embedding and applying a deep learning network named Deep Convolutional Networks. Data analytics plays a very important role in analyzing unstructured data. Structure classification of protein sequence is a hectic task and now deep learning techniques are used to solve this problem. The proposed approach uses NLP-based data representation together with a deep learning approach to efficiently predict motifs.

Chapter 6 - describes a simple minimum distance classifier. The feature extractor results in a classifier performance of 100% recognition rate.

Chapter 7 - focuses on the standard clustering methods like k-means and its variants, comparing them with the proposed technique. Experimental results show that the proposed algorithm is found to work better.

Chapter 8 - considers multi-level grading on significant ripening indices of mango based on microscopic, external and internal attributes using low-cost portable device and sensor technologies, deep learning artificial intelligence and spectral imaging techniques. Such a grading system will help the multiple stakeholders in the mango fruit supply chain deliver quality mangoes to the consumers consistently and cost-effectively.

Chapter 9 - describes efficient algorithms for secure route creation and route failure recovery, which improve effectiveness by reducing communication overhead to give better QOS in mobile ad hoc networks. Secure route creation algorithm using pseudo-random number generation (PRN) to provide an efficient secure route creation is proposed to restrict insider or outsider attacks in MANETs. Data packets may be lost while using time-dependent data delivery (which does not route them to specified destinations).

Chapter 10 - paves the way to efficiently study the signals for echocardiogram, a highly designed system employed to discover heart-related diseases.

Acknowledgments

I extend appreciation to many of our colleagues. I extend my sincere thanks to all experts for giving preparatory comments in the book that will surely motivate the reader to read the topic. I also wish to thank the reviewers who took time to review this book.

I am very much grateful to our family members for their patience, encouragement and understanding. Special thanks are also due to many individuals in Nova Publishers, USA who have contributed their talents and efforts in bringing out this book in the present format.

Suggestions and feedback to improve the text will be highly appreciated.

Chapter 1

Programming a Quantum Computer Using Python

R. Anandan*, Krishnaprasanth B. V. and Sai Sucharitha K.

Department of Computer Science and Engineering,
Vels Institute of Science, Technology and Advanced Studies,
Pallavaram, India

Abstract

Our society may be impacted in ways that we have never imagined by quantum computers. In addition to its many benefits to technology, healthcare, privacy, entertainment, and finance, quantum computing may also extend to other areas of society as well. Although quantum computing is a boon for everyday life, it also carries a few disadvantages, such as potentially breaking existing encryption systems. If organizations are not prepared for the transition from traditional cryptography to post-quantum algorithms, data theft and cyber security issues could occur. Therefore, it is time to learn to program quantum computers. Computing using quantum mechanics is known as quantum computing. There is indeed an interesting combination between computer science, physics, information theory, and mathematics. By controlling microscopic particles (small physical objects) like photons, atoms, electrons, etc., quantum computers are able to compute differently than classical computers. In contrast, Python is an open-source programming language that is widely used for various purposes, such as artificial intelligence, web application development, and data science. Python is a problem-solving language rather than a syntax-based one. Its simplified syntax

* Corresponding Author's Email: anandan.se@gmail.com.

In: Frontiers in Quantum Computing: New Research
Editor: R. Anandan
ISBN: 978-1-68507-816-4
© 2022 Nova Science Publishers, Inc.

gives it an edge over all the other classical programming languages like C, C++, or Java. The purpose of this chapter is to demonstrate how to write computer programs using Python Programming Language for Quantum Computers.

Keywords: quantum mechanics, python, quantum computing, QISKit

1. Introduction

The current generation of computers is smaller, cheaper, faster, more energy-efficient, and even more powerful than the early computers that used to be huge, expensive, and more power-consuming. Architecture, hardware components, and software running on them have all improved to make it possible. Electronic circuits used in computers are getting smaller and smaller day by day. Electronic and electric signals are amplified by transistors, which are small semiconductor devices. These devices are made from silicon. The circuit was made by connecting these transistors into a single silicon surface. The shape of circuits in an IC was printed together in all layers of silicon at the same time. Even when the number of transistors in the circuit is increased, this process still takes the same amount of time. It was the size of silicon, not the number of transistors, that determined the cost of manufacturing ICs. Consequently, the manufacturing and selling of IC and sales of them also increased as a result of the lower price of products. Initially, individual transistors were connected, followed by collections of transistors (Logic Gate) and finally, the collection of Logic Gates became an integrated circuit (IC). Nowadays, a single IC can even integrate small computers onto it.

In 1965, Gordon Moore, a co-founder of Intel, observed that the number of transistors on a silicon microprocessor chip had doubled every year since its inception, while their prices had been cut in half. Moore's Law describes this phenomenon. Moore's Law is significant because it states that computers and their processing power will continue to shrink and speed up over time. Despite the fact that this law is currently putting the brakes on progress in traditional computers, it is not as rapid as it once was [2].

This gives rise to the concept of the tiniest computer, which involves shrinking the circuit down to the size of an atom. However, because electrons inside an atom can become invisible on one side of a barrier and visible on the other, these circuits will not be able to serve as a switch, i.e., they can exist in more than one place at the same time. This is related to the teleporting

phenomenon known as “Quantum Tunneling” in quantum physics. It demonstrates that beyond 5-7 nanometers, the size of the circuits in a traditional computer has hit its limit. The law of classical physics, which gives us only deterministic justification of the Universe, can be used to illustrate the representation and processing of these computers. However, it fails to predict all visible occurrences in nature, which led to the discovery of quantum mechanics, the most significant revolution in physics. As a result, new computing is required that is different from present classical computing in that it stores its state in physical information rather than a circuit. It changes the basic building blocks of a computer that not only expects new type of hardware creation but also a new design, software, and layers of abstraction to facilitate the designers to create and exploit these systems even if they comply. As a result of the quantum phenomenon, there are greater limits on computer architecture, cities scale over time. The design of the hardware components has to be governed by quantum properties [3].

Quantum computing is a novel type of computing based on quantum mechanics that interacts with a probabilistic and unpredictable physical reality. Because quantum mechanics is a more general model of physics than classical mechanics, it leads to a more generic model of computing, quantum computing, which has the ability to tackle problems that classical computing cannot. They employ quantum bits, also known as ‘Qubits,’ to store and modify data, as opposed to traditional computers, which use binary bits 0 and 1 independently to store and manipulate data. Quantum computers are computers that use this sort of computation. Circuits using transistors, logic gates, and Integrated Circuits are not possible in such small computers. As a result, it uses subatomic particles such as atoms, electrons, photons, and ions as bits, together with their spin and state information. As a result, it uses subatomic particles such as atoms, electrons, photons, and ions as bits, together with their spin and state information. They can be stacked on top of one other to create new combinations. They can be stacked on top of one other to create new combinations. As a result, they may execute in parallel and efficiently use memory, making them more powerful. Due to the fact that quantum computing is the only model that can defy the Church-Turing thesis, quantum computers can execute orders of magnitude faster than classical computers.

2. Need for Quantum Computers

Any computational problem that a classical computer can solve can be solved by quantum computers. According to the Church-Turing thesis, classical computers can solve all of the problems that quantum computers can answer. It means they don't provide any additional computability benefits over traditional computers, yet they can't solve some complex and impossible tasks in a reasonable length of time using today's conventional computers [4].

In 1993, Peter Shor demonstrated that quantum computers can help solve these problems much more quickly, in seconds, without being overheated. He devised algorithms for quickly factoring huge numbers. Because they base their calculations on the likelihood of an atom's state before it is known, these have the ability to process massive amounts of data in an exponentially large amount of time. It also explains how a real-world quantum computer may decrypt the secret codes. It has the potential to jeopardize the security of encrypted data and communication. It can expose private and protected secret information. However, the benefits of quantum computers must be considered, as they outnumber their drawbacks. As a result, they are still required, and more study is being conducted in order to ensure a brighter future.

3. Fundamentals of Quantum Computing

The performance of transistors, especially as they get smaller, will be influenced by noise if any type of quantum phenomena occurs, which was taken in mind when constructing the conventional computer. They attempted to totally avoid quantum occurrences in their circuits. However, instead of employing classical bits, the quantum computer uses a different technique and even works with the quantum phenomenon itself. It employs quantum bits, which are comparable to conventional bits and have two quantum states: 0 or 1. However, due to quantum features, it can have both values at the same time, resulting in the concept of superposed bits.

4. Where Does the Concept of Bits Come From?

The communication between transistors within an IC takes place through electric signals. The behavior of the signals is analog. Therefore, their values

are real numbers that change smoothly between 0 and 1. The interaction of these electric signals with the surroundings might result in noise. As a result, a small modification from 0 to 0.1 is required. Therefore, a little change from 0 to 0.1 due to temperature or vibrations from the environment can drastically change the system's behavior. In the environment, there are two forms of noise. The first sort of noise is caused by energy instabilities within the item that occur unexpectedly, such as temperature above absolute zero Kelvin. These are essential. Other types of noise arise as a result of signal interaction. This form of noise could have been avoided if it had been fixed or planned properly. However, neither of these were planned nor rectified at the hardware layer, or perhaps they were left uncorrected on purpose. They are systematic [5].

The IC is created with transistors in such a way that it can function on digital signals (binary bits) instead of analogue signals to overcome these disturbances in analogue circuits. Logic Gates are the name for these circuits. They interpret electric impulses containing real-number values as a binary digit or 'bit' of either 0 (low voltage) or 1 (high voltage) (high voltage). Another form of Gate is a register, which holds a bit or the amount of bits in an input value to be processed further. By limiting the range of values that a signal can contain, gates can reduce noise in a signal. Using logic gates instead of transistors to build an IC simplifies the design by resulting in a powerful circuit that is not susceptible to deflection.

The fundamental unit of quantum information, the quantum bit or qubit, depicts subatomic particles such as atoms, electrons, and other subatomic particles as a computer's memory while their control mechanisms act as a computer's processor. It can have a value of 0 or 1, or both at the same time. It has a million times the processing power of today's most powerful supercomputers. In the world of engineering, producing and managing qubits is a huge task. The quantum computer's processing strength comes from the fact that they have both digital and analogue characteristics. Quantum gates have no noise limit due to their analogue nature, yet their digital nature gives a standard for recovering from this significant flaw [6].

Qubits have two quantum states that are analogous to binary states. The qubit can exist in either state or in a superposed state of both states at the same time. These quantum states have a representation known as Dirac notation.

The state label is preserved between two symbols | and in this notation. As a result, states are expressed as |0 and |1, which have analogue values and both contribute to any value between 0 and 1, given that the sum of the likelihood of each state occurring must be 1. As a result, any quantum bit wave function can be written as a two-state linear combination, each with its own

complex coefficient: |w = x |0 + y |1, where x and y are the coefficients of both states. The state's probability is directly proportional to the square of its coefficient's magnitude. The probability of identifying qubit state 0 is |x|2, while the probability of identifying qubit state 1 is $|y|^2$. These probabilities when summed up must give a total of 1 or say 100% mathematically, i.e., $|x|^2 + |y|^2 = 1$.

5. Properties of Quantum Computing

The quantum item does not exist in a totally defined state in quantum physics. When not detected, it appears to be a particle yet behaves like a wave. Particles' dual nature leads to some fascinating scientific phenomena. A wave function or a sum of possible participating states is used to express the state of any quantum item. Such states are coherent as a result of the constructive or destructive interference of all involved states. The extraction of information from quantum objects when they interact with a broader physical system is the consequence of observation. Quantum measurement is the process of observing quantum objects in this way. Measurement might sometimes lead to the loss of information in the loss of information by disrupting the quantum state. These are some of the properties of quantum objects. Quantum objects referred to here are the qubits in the case of quantum computing. The progress of any quantum system is regulated by Schrodinger's equation that tells us about the change in the wave-function of the system due to the energy environment. This environment is the system Hamiltonian which is a mathematical description of energies experienced from all forces felt by all components of the system. To control any quantum system, there is a need to control this environment by isolating the system from the forces of the universe that cannot be controlled easily and by assigning energy within this isolated area only. A system cannot be completely isolated. However, energy and information exchanges can be minimized. This interaction with the outside environment can lead to loss of coherence and can result in "Decoherence" [7].

The properties are the conceptual rules and mathematical manifestations that describe the behavior of the particles. Quantum computers use three fundamental properties of quantum mechanics to store, represent, and perform operations on data in such a way so that they can compute exponentially faster than any classical computer. The three properties are given as follows [8].

5.1. Superposition

Superposition in quantum mechanics states that any two quantum states can be summed up (superposed) resulting in another valid quantum state. It is a fundamental principle of quantum mechanics. Oppositely we can say that any quantum state is the sum of two or more than two other unique states.

Superposition in quantum computing refers to the ability of a quantum system where quantum particles or qubits can exist in two different positions or say, in multiple states at the same time. It unbelievably provides high-speed parallel processing and is very different from its classical equivalents that have binary constraints. The quantum computer system holds the information that exists in two states simultaneously. Qubits are brought into a superposition by influencing them with the help of lasers so that they can simultaneously store 0 and 1 at the same time. In classical computing, if there are 2 bits, the total possible values after combining we get are 4, out of which only 1 value is possible at any instant. But on the other hand, if there are 2 qubits in the quantum computer. The total possible values after combination are 4 and all are possible at once. It looks like unthinkable Quantum Computing and Communications 6 because it is not like gravity that can be proved easily just by looking at the falling of an apple. The laws of classical physics fail here because superposition only exists in the territory of quantum particles.

For example, when solving a puzzle-like maze, a quantum particle can decide to take the various paths at the same time using superposition. This process matches the function of the parallel computer. Due to this property, the qubit can navigate the maze in exponentially less time than a classical bit.

5.2. Entanglement

Entanglement in quantum mechanics is a physical phenomenon where two or more quantum objects are inherently linked such that the measurement of one rules the possible measurement of another. In other words, a pair or a group of particles interacts or shares spatial locality such that the quantum state of each particle cannot be characterized independently of the other particle's state in the same group even when they are separated by a large distance.

Entanglement is one of the important properties of quantum computing. It refers to the strong correlation existing between two quantum particles (physical properties of systems) or qubits. Qubits are linked together in a perfect instantaneous connection, even if they are isolated at any large

distances such as located at the opposite ends of the Universe. They are entangled or defined concerning each other. The fact is that the state of one particle influences the state of the other. It creates strong communication between qubits. Once they got entangled, they will stay connected even after being separated at any distance. In classical computers, if bits are doubled, computational power also gets doubled. But in the case of Entanglement, adding extra bits to a quantum computer can increase its computational power exponentially. Quantum computer uses this property in a sort of quantum daisy chain.

Some examples of entanglement can be seen in nature such as electrons separated from each other at some distance inside an electron cloud are massively entangled with one another. If one electron is at both the states of spin-up and spin-down with each state having a probability of ½, a similar case is with the other electron.

5.3. Interference

The property of interference in quantum computers is similar to wave interference in classical physics. Wave interference happens when two waves interact with each other in the same medium. It forms a resultant wave with either their amplitudes added together when they are aligned in the same direction known as constructive interference or a resultant wave with their amplitudes canceled out when waves are in opposite direction known as destructive interference. The net wave can be bigger or smaller than the original wave depending on the type of interference. Since all subatomic particles along with light pose dual nature, i.e., particle and wave nature both. The quantum particle may experience interference. If each particle goes through both the slits (Young's double-slit experiment) simultaneously due to superposition, it can cross its path interfering with the path direction. The idea of interference allows us to intentionally bias the content of the qubit towards the needed state. However, it can also result in a quantum computer combining its various computations into one making it more error-prone [9].

6. Python Programming Language

Python is a general-purpose, high-level programming language which is widely used in recent times. Its design philosophy emphasizes code

readability, and its syntax allows programmers to express concepts in fewer lines of code. The most important feature of Python Programming is that it supports multiple programming paradigms, including Object-Oriented Programming, Imperative and Functional Programming, or Procedural styles. It indeed supports a dynamic type system and automatic memory management and has a large and comprehensive standard library. Python Interpreters are available for many operating systems.

7. QISKit

QISKit is an open-source software development kit (SDK) for working with quantum computers at the level of circuits, pulses, and algorithms. It provides tools for creating and manipulating quantum programs and running them on prototype quantum devices on IBM Quantum Experience or simulators on a local computer. It follows the circuit model for universal quantum computation and can be used for any quantum hardware that follows this model [10].

QISKit was founded by IBM Research to allow software development for their cloud quantum computing service, IBM Quantum Experience. Contributions are also made by external supporters, typically from academic institutions [10].

The primary version of Qiskit uses the Python Programming Language. A range of Jupyter notebooks is provided with examples of quantum computing being used. Examples include the source code behind scientific studies that use Qiskit, as well as a set of exercises to help people learn the basics of quantum programming [10].

7.1. Installation

Qiskit can be installed in Python Terminal using the command “pip install risk it.” Note: Installation requires active internet connectivity.

7.2. Importing Qiskit

Qiskit can be imported into a Python program using the following command:

1. import risk it.

7.3. Version

The access the information about the version of Qiskit installed in your system use the following commands

1. from qiskit import __version__
2. import sys
3. print(“Qiskit Version: ,” __version__)
4. print(“Python Version: ,” sys.version)

Output:

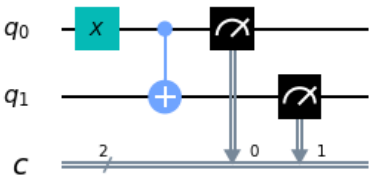
```
Qiskit Version: 0.18.3
Python Version:  3.8.8 (default, Apr 13 2021, 15:08:03) [MSC v.1916
64 bit (AMD64)]
```

7.4. Quantum Circuit

Quantum Computers are essentially different from their core, the “qubit.” Therefore we will be beginning with a bit-level program to build a Quantum Circuit. In this program, we will be creating a quantum circuit with two qubits and two classical bits. First, we apply the NOT gate to the first qubit and then we apply the CNOT gate with the first qubit as the control and the second qubit as the target. Finally, we measure the first and second qubit and record it in the corresponding classical bits, and shall plot the circuit diagram

1. import qiskit as q
2. %matplotlib inline
3. circuit = q.QuantumCircuit(2,2)
4. circuit.x(0)
5. circuit.cx(0,1)
6. circuit.measure([0,1],[0,1])
7. circuit.draw(output=“mpl”)

Output:



7.5. Connecting to IBM Quantum Computing Prototype

Login to Quantum-Computing.ibm.com at https://quantum-computing.ibm.com/ and create an account. After creating an account, click on the account icon at the top right then choose "My Account." From "My Account" obtain your API Token and store it in a text file (in this example:token.txt). Now you can connect with IBM Quantum Computer using the following code:

1. from risk it imports IBMQ
2. IBMQ.save_account(open("token.txt,""r").read())

Output:

```
Credentials already present. Set overwrite=True to overwrite.
```

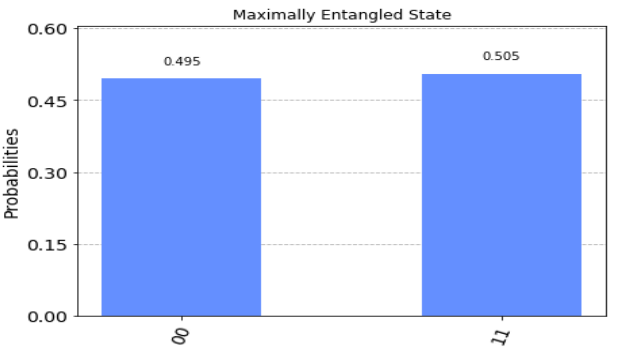
7.6. Executing a Measurement on the IBM Quantum Computing Prototype

Now that we have seen our quantum circuit we need to see our results. This can be accomplished by executing a measurement simulation of the circuit. The execute() function takes the circuit to measure as its first parameter and the simulator that you want to use to do the simulation as the second:

1. from risk, it imports QuantumRegister, ClassicalRegister, QuantumCircuit, Aer, executes
2. from qiskit.visualization import plot_histogram
3. M_simulator=Aer.backends(name='qasm_simulator')[0]
4. qureg=QuantumRegister(2)
5. clreg=ClassicalRegister(2)
6. circuit=QuantumCircuit(qreg,creg)

```
7. circuit.h(0)
8. circuit.cx(0,1)
9. circuit.measure(0,0)
10. circuit.measure(1,1)
11. job=execute(circuit,M_simulator)
12. hist=job.result().get_counts()
13. plot_histogram(data=hist,title="Maximally Entangled State")
```

Output:



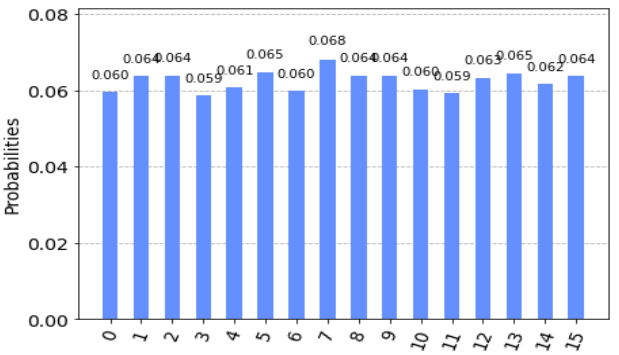
7.7. Random Number Generation using IBM Quantum Computer

Do you think the python random function returns a random number all the time? No. In fact, in classical computing, it is very hard to randomly generate numbers. The random() function is a pseudorandom number generator. It produces an output by taking a definite input and processing it in a deterministic way. This program resolves pseudo-randomness using quantum superpositions. Because we know that the state of a qubit remains unpredictable until we measure or observe it.

```
1. from qiskit import QuantumRegister, ClassicalRegister,
2. QuantumCircuit, BasicAer, execute
3. from qiskit.visualization import plot_histogram
4. qreg=QuantumRegister(n)
5. creg=ClassicalRegister(n)
6. qc=QuantumCircuit(qreg,creg)
7. for j in range(n)
```

```
8.  circuit.h(qreg[j])
9.  qc.measure(qreg,creg)
10. job=execute(qc,BasicAer.get_backend('qasm_simulator'),shots=819
    2)
11. bit_counts=job.result().get_counts()
12. int_counts={}
13. for bitstring in bit_counts:
14. int_counts[int(bitstring,2)]=bit_counts[bitstring]
15. plot_histogram(int_counts).
```

Output:



8. Major Challenges in Quantum Computing

The good news is that at any instant of time, the quantum state with the same number of quantum bits can stretch over all possible states as compared to classical computers and thus works in an exponentially massive space. However, to be able to use this space requires all qubits to remain interconnected. Even after such progress, improvements are still needed. The bad news is that making new and high-quality qubits does not guarantee the creation and efficient use of fault-tolerant quantum computers and is still having challenges in its path [11].

Qubits cannot naturally ignore the noise. Hence, the quantum system is more error-prone. It suffers from decoherence. The biggest challenge is how it can handle any undesirable deviations or noise in quantum computers. Classical computers can produce clean noise-free outcomes by simply putting

their state as off or '0,' which is not possible for quantum computers where errors occur in physical circuits. Qubits will gradually lose their information as well as interconnection (entanglement) between each other. The error rate is seen as a design parameter for such systems which should be improved in large qubit systems also. However, to make the qubits stable and error-free, they are being insulated from the outside environment in super-refrigerated fridges or vacuum chambers and accurately handled [12].

Qubits are neither completely binary nor digital. It is having analog properties also. Gate can reject noise by dealing with the input signal value of 0.8 and treating it as 1. But in the analog signal, every value between 0 and 1 is permitted since they have their meanings. Signals cannot be checked for any kind of noise or corruption, since 0.8 can be 1 with some error or 0.8 without error. Presuming the error as 0 like Gates does or taking some noise value even if it was not present there can affect the adherence of the resulting quantum computation. Hence, there is a need for algorithms like quantum error correction similar to the logical error correction in classical computers. These algorithms can be run on a noisy gate-based quantum computer to eliminate the errors and noises present in them [13].

It is possible to employ a Quantum Error Correction algorithm on a quantum system. But quantum error correction requires dealing with the overhead such as a large number of qubits and their fundamental operations and generally needs more resources. Also, problems with large data inputs require a large amount of time to create the input quantum state that would monopolize the computation time lessening the quantum benefits.

Quantum algorithm development is another challenge since achieving quantum speedup expects entirely new types of algorithm design as the speed of computation depends on the design of the algorithm. The design of the algorithm should be corresponding to the number of qubits used. Further development of software tools in addition to hardware, is required to create and debug quantum systems to help explain unknown issues and push towards designs.

Debugging quantum hardware and software is of utmost importance which depends on memory and intermediate machine states in classical computers. But in the case of quantum computing, states cannot be copied directly for later evaluation, and directly measuring intermediate states can bring it to halt. Hence, new strategies for debugging are essential for their development.

Conclusion and Future Scope

A significant amount of struggle is remaining before a practical quantum computer can be launched. Some future advancements are needed. Some of the future needs are enabling a Quantum Error Correction algorithm that requires low overhead and decreases the error rates in qubits, developing more algorithms with lesser qubits for solving problems, reducing circuit thickness so that NISQ computers can be operated, the advancement of methods which can verify, debug, and simulate the quantum computers, scaling the number of qubits per processor in such a way so that error rate is maintained or can be improved if possible, interleaving of operations in a qubit, recognizing more algorithms that can reduce the computation time and creating input-output for the quantum processor.

Such 'Quantum games' are predicted in the future that will give unexpected situations and results that a player can experience because quantum computers will take all the possible operations and throws them into the game randomly due to their quantum properties like super positioning and entanglement of qubits. It will be a never-ending experience.

Quantum computing in Cloud has the potential to overtake business initiatives like in other emerging technologies such as cryptography and artificial Intelligence. Since the classical simulation of fifty qubits is equal to the memory of one Petabyte that doubles with every single qubit added [14], the memory required should also be large enough to provide an environment for application development and testing for multiple developers to simulate quantum computers using suitable shared resources.

AI and machine learning problems could be solved in a practical amount of time that can be reduced from hundreds of thousands of years to seconds. Several quantum algorithms have been developed such as Grover's algorithm for searching and Shor's algorithm for factoring large numbers. More quantum algorithms are coming soon. Google has also declared that it would produce a workable quantum computer in the following 5 years with a 50-qubit quantum computer and will achieve quantum supremacy. IBM is also offering commercial quantum computers soon.

The progress of development in the field of quantum computers depends on many factors. Interest and financial support from the private sector can help develop commercial applications for NISQ computers. It depends on the progress of quantum algorithm development, the availability of enough investment in the quantum technology field from the government, and the exchange of ideas among researchers, scientists, and engineers [15]. To

illuminate the limitations of quantum technology, a defensive result is also beneficial. It can help in overcoming those negative results which can lead to a discovery.

References

[1] Copeland, B. J. (2000). *The modern history of computing*, https://plato.standford.edu/entries/computing-history.

[2] Thesis, T. N., & Wong, H. S. P. (2017). The end of Moore's law: A new beginning for information technology. *Computing in Science & Engineering*, 19(2), 41-50.

[3] Feynman, R. P. "Simulating physics with computers (1982)," *International Journal of Theoretical Physics*, Vol. 21, Nos. 6/7.

[4] Arute, F., Arya, K., Babbush, R., Bacon, D., Bardin, J. C., Barends, R., & Burkett., B. (2019). Quantum supremacy uses a programmable superconducting processor. *Nature*, 574(7779), 505-510.

[5] Grumbling, E. and Mark Horowitz (2019), "2 Quantum Computing: A New Paradigm. "National Academies of Science, Engineering, and Medicine. *Quantum Computing: Progress and Prospects*. Washington, DC: The National Academies Press. doi: https://10.17226/25196.

[6] Benett, C. H. and David P. DiVincenzo (March, 2000), "Quantum Information and Computation," *Nature*, Vo. 404, 16.

[7] Amin, M. H. S., D. V. Averin and J. A. Nesteroff, 2009, Decoherence in adiabatic quantum computation, *Physical Review* A 79(2):022107.

[8] Amyx, S. (2017), "*quantum-computing-series-part-4-superposition-in-quantum-mechanics-381b98180f62*," https://medium.com/@ScottAmyx/quantum-computing-series-part-4-superposition-in-quantum-mechanics-381b98180f62.

[9] Rose, M. (2011, "Quantum Interference" *WhatIs.com*, TechTarget https://whatis.techtarget.com/definition/quantum-interference.

[10] Wootonjames (2018), Qiskit, *Wikipedia*, En.Wikipedia.org, https://en.wikipedia.org/wiki/Qiskit.

[11] Franklin, D., Chong, Frederic T. (2004). "Challenges in Reliable Quantum Computing." *Nano, Quantum and Molecular Computing.* Pp. 247-266. doi: https://10.1007/1-4020-8068-9_8. ISBN 1-4020-8067-0.

[12] Joseph, M., K. Elleithy and M. Mohamed, "A new Quantum Processor Architecture," *2019 IEEE 10th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON)*, New York City, NY, USA, 2019, pp. 0483-0487. doi: https://10.1109/UEMCON47517.2019.8992935.

[13] Kandala, A., K. Temme, A. D. Corcoles, A. mezzacapo, J. M. Chow, and J. M. Gambetta, 2018, *"Extending the Computational reach of a Noisy Superconducting Quantum Processor,"* arXiv:1805.04492.

[14] "Multiple Quibits (2017), "*Microsoft Quantum docs*" https://docs.microsoft.com/en-s/quantum/concepts/multiple-quibits.

[15] Office of Science and Technology Policy, 2018, *National Strategic Overview for Quantum Information Science,* https://www.whitehouse.gov.in/wp-content/uploads/2018/09/National-Strategic-Overview-for-Quantum-Information-Science.pdf.

Chapter 2

Blockchain-Based Quantum Key Distribution Approach

M. Karthik* and R. Balakrishna

Department of Computer Science and Engineering,
Vels Institute of Science Technology and Advanced Studies (VISTAS),
Tamil Nadu, India

Abstract

Quantum computing uses the aggregation and connectivity characteristics of quantum mechanics to calculate. Supercomputers are often used to address issues that are incapable of solving with computer devices in block chain technology. Quantum circuits, often known as qubits or quantum bits, are perhaps the most frequently used quantum structures. Quantum cryptosystem is more secure than traditional approaches in crypto. For quantum computation, Shor's technique and Grover's approach are the most often utilised approaches. Rectilinear or Radial bases are used in a chaotic way for encoding and decoding. The transmission of electronic keys QKD is a symmetrical access control technique for cryptography.

Authenticity and secrecy are two key features of QKD. To enable quantum secure authentication scheme and significant measures communications in the networking, a public key mechanism and a homogeneous decryption key are employed.

The Morse code is utilised for encryption process inside the current proposal. The Morse code is a method of encrypting bits as visible light, which is necessary for the dissemination of quantum keys in block chain technology.

* Corresponding Author's Email: karthikcpta@gmail.com.

In: Frontiers in Quantum Computing: New Research
Editor: R. Anandan
ISBN: 978-1-68507-816-4

© 2022 Nova Science Publishers, Inc.

Keywords: Qubits, QKD, quantum cryptosystem, block chain, validation and privacy

1. Introduction

Privacy and encryption are critical for modern day-to-day connectivity via numerous networks. Conventional cryptographic methods for conventional networks are vulnerable to a variety of assaults which can be countered by quantum mechanics, and they require complicated mathematical calculations. The QKD technology has no complicated calculations and offers quantum cybersecurity. QKD is mostly used to locate network eavesdroppers. The transmitter and receiver exchange a pair of keys across the created quantum channel. QKD algorithms vary in terms of modulation schemes, cryptography methods, as well as the quantum channels are formed. The initial QKD system, designated as DVQKD, employs discrete variables and converts them utilizing light polarisation. This technique was then refined into the BB84 technique. Because qubits contain four alternative representations, separate state polarisation was utilised to cipher the fragments in the BB84 algorithm. The orthogonal or rectangular shapes foundations are used in those qubit forms. In encryption algorithm, the foundations that evaluate light intensities are randomly selected. Most of communication across a network takes place by means of a communication medium. The basis is agreed upon by both the transmitter and the recipient. The pieces with various bases are erased, as well as the password is constructed and validated for accuracy, ensuring that no data is lost. The BB84 technique is similar to the B92 technique, only it only has two components rather than four. The eavesdropping does have some foreknowledge on how to listen in on message being relayed.

Entrapment of data is an additional approach for generating secrets. The message transmitted mostly from the transmitter to the recipient in a single-way CVQKD technique, however in a double-way CVQKD methodology, the recipient begins interaction after receipt of information, and the transmitter provides hidden as well as the received signal. The dissemination of quantum keys QKD is used to send a secure key between two nodes situated in various corners of the world. Improved security is provided by the QKD than existing methodologies.

Security concerns to business information management must be protected before any IT operation can be completed, particularly Supervision Control and Data Acquisition (SCADA) platforms [1]. SCADA methodology is to

collect monitoring information and transfer instructions to controllers in wireless sensor networks (WSN) environment [2]. The majority of these technologies really aren't immediately exposed to the Network, but instead to corporate management and technology platforms, as well as connection between producers and contractors via public devices. The crucial SCADA platform's separation will put it at danger of needing to acquire signals from the environment.

The implementation of such procedures leads to new types of data security breaches that are more hard to regulate [3]. For instance, its Stuxnet virus got past APCS gates while using different techniques like Usb flash drives or CDs, or by employing standards that defenders had set to ignore. As a result, the goal of SCADA security protocols is to make efforts that makes the monitoring system, connections, equipment, and tuples more resistant [4].

The recent trend in constructing a transportation infrastructure for SCADA gathering data dictates the employment of wirelessly conscience networks with properties such as cluster equalization, rapidly adjust architecture, and conscience after a failure.

The IEEE 802.11 [5] offers an excellent framework for building a trustworthy WSN for gathering data. WSNs are progressively displacing routers that are used in processes that control advanced technologies [6], inside the mortgage market to regulate thermal production [7], bright lights, wind, and airflow [8], in advertising water usage accounting, in fire surveillance systems, remote control, telemedicine, and other applications.

Professionals employ tools and solutions for a cross defensive architecture for the elements of the SCADA systems and sensitive information communication with common data communication to defend against various forms of attacks.

1.1. Key Organization for Information Encryption in WSN Environment

The encoding of asymmetric cryptography analysis is required on the use of a secret cryptography or an asymmetrical cryptographic signature. For authenticity of SCADA process components and sensing networking devices, unique symbols are utilised, and hash values have been used to verify the authenticity of sent information. There really is a lot of private documents in every efficient information technology that needs to be protected against

compromise, which necessitates the use of effective management technologies.

The methods for secure generation, key use, and administration are difficult tasks even if cryptosystems to prevent breaches are properly designed. Two issues are listed below:

- How can you make credentials with mathematical qualities you need?
- How can the credentials again for considered a challenge in WSN be sent securely to the respondents?

Filled with self, accidental interconnections throughout navigation, as well as the sheer randomness of communication exchanges, the complication of access control in wireless communication is defined by the lack of any set pathways metadata. The fundamental goal of key generation is to offer vital information to members is considered as a challenge task in wireless networks in order to perform private information sharing through a transport layer security. The following are the key-management methods that really should be applied in the controller:

1. System components are registered as conversation members.
2. The creation of cryptographic hash keys.
3. Key transfer and dissemination among components of a SCADA platform's WSN.
4. Handling the transfer partners' and keys' relationships.
5. Keys must be replaced.
6. Key restoration in the event that it is accidentally destroyed.
7. The destroying of keys, either intentionally or unintentionally.

1.2. The Extensive Keys Construction of the WSN IEEE 802.1 Equipment

The encrypting and preservation of significant information when creating confidence amongst homologues is the secrecy method in the Sensor 802.1 specifications, including at the point of key deployment and during the data exchange procedure [9]. The IEEE 802.15.4 specification governs the authentication scheme, which uses unique characteristics to offer protection. Specs Cryptography is supported by the 802.1 Pro Extracted Features, which also defines relative importance of different transmission and cryptography.

While sharing data just at command line, an extra cryptography technique can be employed which can be decoded by every other network interface, regardless of the fact that because they have same connection key.

The safety structure is predicated just on AES homogeneous encryption method with 128-bit keys, which could be connected with both the networking (network key – NK) or route (link key – LK) in line only with 802.1 standard. The keycard (Pk), essentially ensures their conformance, is used in the key generation. A protected environment is necessary to receive the primary keycard.

The Wireless router is responsible for the transport of central management keys, which many other stations recognize. To obtain the NK and transaction keys for something like the LK for communication, every device in the network should contains details of central management keys. The central management key empowers any connectivity for devices in communication network or reorganization, i.e., while dealing with access control lists (ACLs). The central controller is usually not only for WSN's coordination, and it might be linked to a command centre. Different kinds of keys are used in the WSN 802.1 guidelines:

1. A primary key, that is used to generate the connection bridge key as an original shared hidden message across two devices.
2. Every network device will have its own network key (NK), which offers privacy there at network topology. While detaching and restoring devices towards the system, those keys have been used. The system key might well be updated by the centre on a regular basis and transmitted to everyone terminals inside the key pair.
3. There at applications level, transaction LKs offer secured unicast transfer of packets across stations.

Because 802.1 protection is dependent on cryptographic cryptography, the data packet transmitter and receiver should possess this very same common key. For such users of the communication environment, there seem to be different techniques of consignee: (i) pre-installation, (ii) communication from access control unit, and (iii) key composition even by conversation parties. In the accordance with pre, these keys were entered in the terminals or PLC before the computer's software is updated. The keys are sent to the machines by the strategic control unit in the second scenario. Inside the third scenario, another of the parties produces its very own keys then passes them to the other without sharing information.

The parties are sent another secret key to cryptographic mechanism through the asymmetrical transmission, which creates the various challenges:

the requirement for authenticated key communication to every customer over a private or secret connection; the difficulty of identity management, which entails a polynomial rise in the number of values to produce, transport, preserve, and destruct for every adjacent nodes inside the wireless channel.

In communication links, an advanced encryption strategy with cryptographic key is employed to tackle these challenges. Through use of encryption methods solves the issue of key exchange inside the network, but also presents the issue of authenticating the obtained keys including their originating authenticity, which is particularly important in wireless networks. Whenever a communication is originally submitted to the hashing, the secret key entity may ensure the integrity of the receiver signature through using decryption key technique as an electronically electronic signature. The RSA cryptography employs a strategy of consultative approach of AES cryptography in which the originator is first included to the statement's unique identifier.

Even though this technique eliminates the difficulty of homogeneous transactions involving the first transmission of the secret to another participant, similar devices are costly in relation to key duration, bandwidth, and entire system efficiency, making it unsuitable for sensing devices. As a result, the mixed cryptosystem with WSN is more appealing since it integrates the benefits of an asynchronous system with capabilities of homogeneous crypto algorithms.

A particular form of key (unsymmetrical key - AK) should be used to authenticate a symmetrical mutual authentication relationship, in addition to these core configuration used throughout the Wifi specification.

The MK, NKs, and specialized AK keys have such a lengthy existence, whereas transaction keys get a finite lifespan and save resources, WSN may adopt essential information monitoring system and a private key mechanism inside the packet forwarding employing packets of navigation parameters and reception validation.

2. Related Works

A BB84 [10] quantum key exchange technique was invented in 1984 by Charles Bennett and Gilles Brassard. The transmitter produces two randomized characters, which the recipient examines. Qubits confirm for

comparability during the computation. If just under a tolerable level of disagreement exists, the quantum link is correct, as well as the secret keys bit could be obtained by using remnants. Through using QKG-AKA [11] technique, proposed security procedures for effective safeguarded key exchange. As a result, the likelihood of multiple assailants assaulting is reduced. Energy QKD [12] technique is used to provide an intrusion prevention strategy. Simultaneously, for dispatching automate, distribution mechanization, electrical information gathering, and teleconferencing, a current breed of QKD related information privacy transportation infrastructure has been created. In a multi-party setting, creating secured connections requires basic requirements [13]. If any computing constraint makes the secret setup impossible, this can only be accomplished with quantum theory. The conversation can simply be intercepted and captured. However, to prevent the harm caused by algebraic and computationally private group secrets, anonymity must be preserved. The AVISPA programme was created to combat bogus access point attacks [14] using AKA method for high-speed data transmission. Reliability and accessibility are all important benefits of a 5G network. As so many equipments are linked to the environment that results congestion in the system. To lessen the jamming effect, uses a Dual – Homed Switching Network (DSN) which combines with Software - defined networking Networking (SDN) and Network Function Virtualization (NFV) [15]. DSN offers an improved security mechanism for any and all wireless network technologies.

To enhance authenticity in Mobile Edge Computing (MEC) services by offering the Extensible Authentication Protocol authentication and key agreement (EAP - AKA) [16] interface, that employs a persistent identity single way and hash technique. The confidentiality [8] of the location of SDN and NFV has to be preserved. For effective and rapid changeover identification in the application layer, the Kolmogorov – Smirnov (K-S) assumption test [17] is proposed. With the 5G network and the wireless technology, a new period based on the Internet of Things IoT begins. For encrypted transmission in Iot systems, authentication scheme management is necessary. The ITEF [18] assesses the trustworthiness of 5G networks and makes recommendations for improving protection.

The conventional electric system establishes a vulnerability management power to enhance reliability by employing surveillance systems including routers, intruder identification techniques [19], and cryptography equipment to assure safe information exchange to a certain extent. The key generated in cryptosystem are refreshed periodically to increase security mechanism in

network. However, the rate of upgrading is weak, lowering high bandwidth protection. The device's dependability and responsiveness are inadequate, and its secure effectiveness is even further affected, in respect of the concealed threat of becoming stolen through conventional main online communication. In [20] it describes a double parameter authentication scheme control system. Information privacy and sharing identification are 2 aspects that can survive a variety of assaults.

Due to constant changeover, connection delay increases, which would be mitigated by employing balanced stable data to facilitate a rapid security mechanism applying the Neyman Pearson (NP) assumption test [21] that improves accuracy and reduces significant delay. In conjunction with European Training Network (ETN), we pledged to develop a safe networking classification scheme by lowering the power consumption in tiny cells [22]. Cellular modems are used to substitute the cellphone with tiny cells. The customers' information is held private so it has minimal security features. The main difficulty is leading to significant amongst network users. An unsecure wirelessly media is used to create a common key. Low Density Parity Check (LDPC) techniques were used to guarantee secrecy [23]. In the existence of an unauthorized person, this technology performs. Portable network elements with network coding (NC-MSCs) are being used in 5G technology to allow high bandwidth rate equipment connection. The authors propose a Distributed Key Management Strategy [24], that delegates authentication and authorization functions via threshold secure communication. Each cluster has primary secret keys that provide the cluster with a certificate which could be used everywhere, any time in environment. The secure context information (SCI) factor is utilised in the systems to enable system performance by lowering delay and improving consistency [25]. In HeNet, it can be used for quick validation. The traditional network's turnover authorization approaches cause a lot of delays and complexities. In [26], the authorization changeover content is collected based on the destination address, and SCI is retrieved from features of the radio medium being used for interaction between the client and the gateway. The competence-based confidentiality transference validating system has achieved common verification, and UE and BS successfully reached the highest agreement, which significantly decreases the price of turnover.

The security of the system is a crucial element that should be preserved. In the LTE architecture, the evolving transmission subsystem authenticity and key agreement (EPS-AKA) structures are significant security mechanisms [27]. The EPS-AKA mechanism is regarded as a critical and long-term

alternative for LTE privacy, even though it has the disadvantage of employing a single persistent key to generate all future keys. The EPS-AKA is a method for overcoming authorization delays and reducing network latency through obtaining authorization from external systems, which lowers operational costs. Switchover and key distribution amongst participants are less secure in LTE connectivity. The elliptic curve cryptography technique premised intermediary authentication [28] is often used to enable safe changeover at a low operating complexity. It will serve as a foundation for the rest of article's advances, and a benchmark against which effectiveness may be measured. Assume there are two endpoints. Bob (B) and Alice (A) want to compromise on a key. In the existence of a radio rayleigh fading connecting both Eve and the eavesdropper (E). Either Alice or Bob have the ability to connect by a wireless connection. They can also transmit data as is commonly thought inside the documentation, through using community connection [29].

For exchanging diversified hidden photos, [30] presented portion susceptibility, various size asymmetric shared restoration technique and complete sensibility various sized symmetrically communication techniques. These two approaches use a similar feature to exchange these hidden visual effects lights, which are used to restore the hidden pictures using the distributed dark spots. In the Diffraction sector, [31] suggested new picture cryptographic techniques based on row scanner compressed phantom photography, (t, n) cutoff steganography, and phasing recuperation. Following wavelet, the decryption matrix was initially identified using bucket detection in the encoding process.

The prediction key used in column checking compression phantom photography was partitioned and distributed into two pairs of sub-keys, which would then be reconstructed using two - stage masking (POM) keys using predetermined image pixels using a (t, n) minimum selective encryption procedure. In [32], the ubiquitous transfer was used to create a robust and efficient number system and multisecret photo sharing mechanism. The ubiquitous transfer valuation mechanism is used in digital mathematical computation techniques, whereas the decryption of secret data improves the irrationality feature of ownership and fulfills the cutoff characteristic with an effective completing approach. It suggests a (t, k, n)-Essential Hidden Information architecture for image pixels and over nonnegative GF(pm) that is reliable and therefore does not involve additional pretreatment. Nonetheless, the approach covering GF(pm) having pm > 28 remained least lossless, so it worked properly whether the number of participants was greater beyond 255. This technique, in comparison to GF(28), was completely uncompressed and

reduced sharing capacity, and it had no problems like differential darkness sizes, sub-shadow connections, or the use of differential coefficients, among others.

In [33] proved that (n, n) secret systems and similar approaches fail to meet the security requirement. Having data on (n 1) or fewer sharing, for instance, essential information about the hidden picture might be gathered using expert computations. The researchers have solved the critical vulnerability with an established version that generates permutation sharing and outperforms control scheme Boolean related solutions in terms of processing time. In [34], irreversible data hidden in scrambled photographs was subjected to image encryption. A picture cryptography based on Shamir's visual cryptography was described at the time. Finally, two essential strategies based on differential growth and differential logarithmic flipping, or certain modifications, were presented. In [35] presented a selective encryption approach with essential shadowing that is flexible (t, s, k, n). In this strategy, at least k reflections that comprise at best t primary lights can gradually reconstruct the hidden information; while all essential lights are incorporated, the entire hidden message can be reconstructed. This solution combines the two advantages of secret picture transmission and the scalability of hidden file sharing via fundamental lights, which was previously possible in a multitude of scenarios. To recommend three (k, n), [36] constructed three logical activities: bit-level XOR, coverage functional (COV) (1, 7, 3) from (7, 4) Hamming protocol, and COV (2, 8, 4) from (8, 4) shortened Hashing algorithm. k to n picture darkness can decode arbitrary significant regions here on entire image then continuously recreate the picture to use these approaches.

In [11] suggested a curvelet hidden social sharing strategy that uses the optimum Homomorphic Encryption (HE) approach to encapsulate darkness photographs. To construct post, the hidden data must be transformed to use the Discrete Wavelet Transform (DWT). Different lights are created, decrypted, then deciphered for every darkness using this process. The encrypted light was found by selecting a small number of such 'n' lights that were translucent and stacked on top of each other. Every shade is demodulated using the HE approach to enhance shadowing protection.

The novel Oppositional oriented Harmonic Searching (OHS) technique was utilised to create the optimal key in order to increase accuracy. The author of [16] established the FDSR and PDSR techniques for distributing hidden photos of various sizes. Entities could, nevertheless, be impacted by distributed shadowing randomization. In [18], ubiquitous sharing was used to create efficient and robust computation through a photo sharing mechanism.

This is not, unfortunately, ideal with photos of a complicated dimension. [19] developed a (t, k, n)-ESIS architecture for image pixels over input space GF(pm) that is secure and will not involve additional processing. This strategy fails to produce a secured communication platform that is efficient. In [37] proved that (n, n)- secret media sharing approaches fail to meet the security measure. They remain, unfortunately, harmed as a result of shared shadowing unpredictability and limited latency. For recoverable communication hidden in protected photos, [38] developed steganography. It has the highest cryptography effectiveness Slowing the randomized key could slow it down. Confidentiality is present in a large percentage of photographs used in many, making a world of difference. This private information is frequently misused by hackers for illicit purposes such as threats, cyber warfare, cyberattacks, extortion, and bank fraud. Considering such potential risks, it seems that developing security precautions is critical to safeguarding confidential materials. A few new incidents have exposed the flaws and weaknesses of optimizers. As a result, unique cryptosystems with exceptional traits and capabilities must be devised to address this problem. Secrecy sharing (SS) systems are seen as one of the options for getting around established encryption techniques. Subscribers of an SS processing tasks section of the information, which must satisfy two requirements: (1) Confidentiality: no information pertaining the hidden is picked up by any disapproved portion of participants; (2) Recoverability: every authorized subgroup of participants can retrieve the secrets.

Interacting equipment can employ preloaded keys to safeguard their interactions thanks to which was before encryption schemes. Keys are preloaded into instrument storage after distribution in such approaches [39]. The identical insertion pathway is followed by everyone's key pre-distribution procedures. The manufacture and distribution of keys is the first phase. When monitoring a system for a system, a key is created and stored inside the device's storage. The discovery of a shared key is indeed the second phase. Each end looks through respective memory with at least one specific key to authorize the relationship at about this point. Ultimately, the shared key is used to create a secure connection. The current system was distinguished from one another by dominance in one or even more factors. Equipment cost, computational burden, regional connections, worldwide connections, and resilience are the variables in question.

In memory and computational, processing, and power consumption, information in such a way techniques offer the advantage of reducing that complexity incurred by key thing generation technologies. Many constraints

limit the use of information in such a way in WBAN. To begin, there are squandered capabilities, just like RAM used to hold keys in every sensing element. Consequently, communication parties may be unable to relate to the topic keys. Information in such a way, techniques include the pseudorandom which was before technique, the composites pseudorandom which was before framework, as well as the spread spectrum retraining method.

Safety keys based on the channel attributes were initially suggested to install public-key methods in cybersecurity. These are physically security protection approaches that produce a private key by exploiting transmission randomization features. Received Signal Strength (RSS), Channel State Information (CSI), and Commitment Data Throughput are the most commonly utilized network parameters (CIR). The public key process could be subdivided into 4 phases depending on the input attributes: Information resolution, and confidentiality augmentation, network monitoring, and quantification. Information, which is network detection, that involves simultaneously transmitting devices measuring a stochastic property of the medium, like the RSS dataset. Gray encoding could be used to decrease key discrepancy in cross compression [40]. Data reconciliation procedures are used to delete or repair key bits or inconsistencies just at opposite edges during the reconciliation phase. Lastly, enhancement of confidentiality is conducted. By raising key unpredictability and dissembling any insufficient information an intruder may well have collected, anonymity augmentation strives to enhance the corresponding keys.

In practice, determining the quantity of knowledge that has been released is difficult. This technique is capable of reducing the latency imposed on the sensing when identifying basic patterns and generating an absolute key. Unfortunately, at an expensive cost, this method is not practical in WBAN. In [41] would be the one to propose the use of physiologically single carrier key establishment techniques to safeguard device interactions. It mentioned a few basic premises.

To begin, bidirectional communication devices should concurrently capture the very same kind of bit stream. Secondly, the key creation procedure is based on different classifiers from tools for evaluating. Furthermore, if the transmitting devices share certain characteristics, typically consists could be obtained. Because of the changing aspects of the human body, physiological data recorded by devices in various body regions show comparable patterns instead of comparable results. To combat all this, imprecise crypt systems like PSKA [42] and OPFKA [43] have been developed. The main idea behind a fuzzy crypt is to hide valid qualities that constitute the key's foundation by

introducing noise data. The repository is a matrix containing genuine characteristics as well as noise data.

3. Proposed Methodology

The qubits are utilised to create a quantum interface. Such qubits can be in either a 1 or a 0 position. Qubits are represented as diagonal lines and upward/downward lines in two distinct ways. The message is stored as qubits and sent over a quantum link. The qubits are decoded in a random order at the recipient. Between both the transmitter and the recipient, a Quantum Key Converter was employed. The distribution verifies that the originator and recipient messages are identical. At least 50% of the encoded and decoded text must coincide. Channels are 50 percent synchronized, the network is designated as a quantum channel, which is utilised as a secure will be in the foreseeable. If the mismatch becomes less is below 50%, the third coalition's connection has already halted, and the data has been altered, requiring the connection to be altered for future initiatives [22].

As seen in Figure 1, the server and the user are the 2 elements that make up the Transmitter and Recipient connection. Clients send queries, which the server answers. Both communicating parties tend to communicate over a network connection, which may or may not be on the same machine.

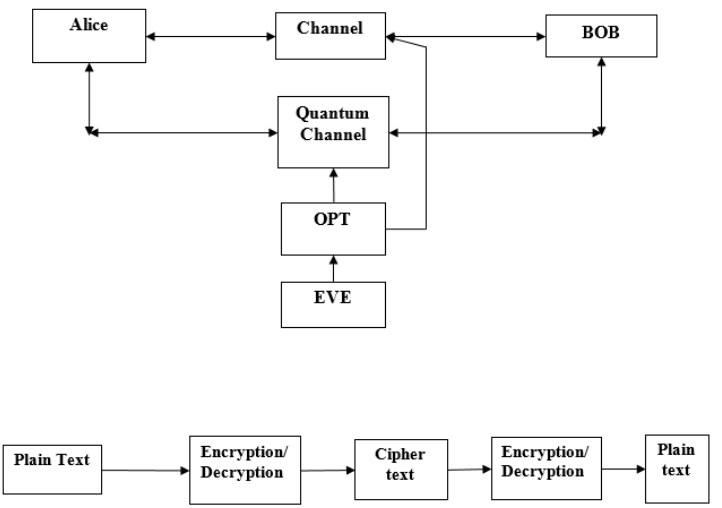


Figure 1. QKD structural design.

1. The socket.gethostname() method returns the present host name, is utilized to run interpreter in Python. This method can be applied in grouping with socket.gethostbyname() to get the local IP address of host's.
2. A server's bind () function connects it to a specific IP address and port address, permitting it to wait for user requests on such IP address and port address.
3. Listen () is a server function that turns it into a listener. This allows the server to keep track of connection requests and reply appropriately.
4. Accept () is a server technique that provides the customer's domain and even an established relationship between certain sender and the receiver.
5. Whenever a client's conversation is complete, the relationship must be tidied up utilizing close ().

3.1. Sender

In order to define a web server, we must first initialise Socket module. A 5000 port is used to communicate in between user and the server in this case. Some other port can also be used. This user is being awaited by the accept() method. The Socket object is provided when users link to the necessary specific port.

3.2. Receiver

In order to construct client software, we must first create objects of the Socket object. The IP address or domain of the host, as well as a destination port, must be entered there.

Figure 2 shows Alice sending a request packet to Bob via the connection. Bob's answer to Alice's request includes a check of the parties to reach in able to link for conversation. These processes take place in the traditional route. If the device is connected effectively with no lost data, the quantum phase is formed.

For encrypted communication, a quantum network is utilised. Alice and Bob have cryptographic keys that they have used as a private key for communication purpose.

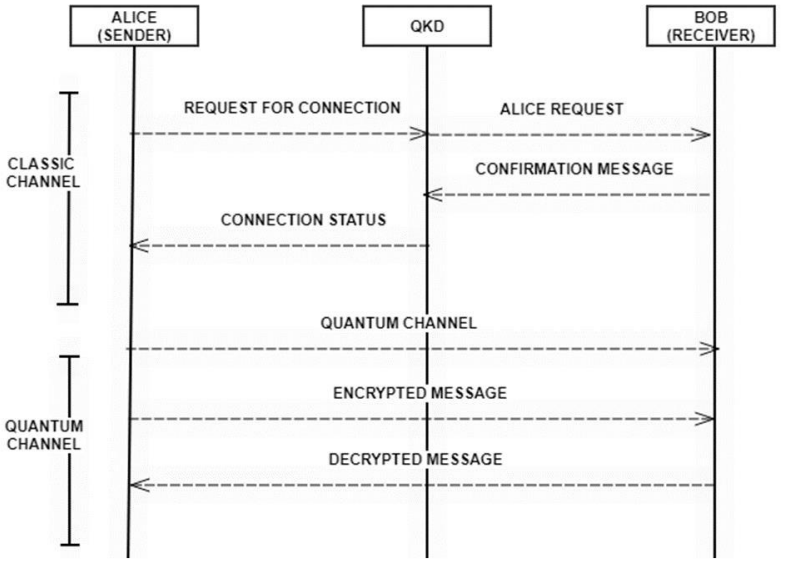


Figure 2. Flow Graphic Representation.

3.3. Data Encryption

The document's authenticity is encrypted as a bit sequence. Transmitter utilised Light signals to encrypt messages.

Recipient encrypted as such because he could just communicate decimals and ones.

1. One is just a small marking, such as a dot(.) or "dit," another is a longer symbol, such as a dash(-) or "dah."
2. There seems to be a single dot period or one meter log interval every after dit or dah.
3. There is indeed a small gap among each word.
4. There is indeed a moderate space among each phrase.
5. Whenever decoding is switched to decimal, the brief mark (dot) becomes 1 and the extended sign (dash) becomes 111.
6. The interpersonal and inter space among characters is denoted by the number 0.
7. The small interval is denoted by 00, whereas the middle distance is denoted by 000.

3.4. BB84 Procedure

The BB84 method's steps are followed: The transmitter selects the information as just a bit, k, and one pseudorandom number sequence, b, in the first stage. The qubit is encoded inside the regular scale by the transmitter. It will become more apparent as each foundation is represented by two horizontal lines, with only two distinctive foundations inverted by 45 degrees. As a result, the encoding of every qubit was as follows: D-basis stands for transverse premise, while R-basis stands for curvilinear basis. The transmitter transfers the qubits to the recipient following encrypting the ‘n' qubits. The transmitter also creates a randomised binary string bi, which contains ‘n' bits and defines whether the base will be used for observations. The results of the observations ki, as well as the accompanying fundamental bits bi, are stored by the reception. The transmitter and recipient then match their b and bi core bits. Since bi is not identical to b, the recipient's qubit was encrypted with a likelihood of 1/2 on a separate scale to that of the sender's. As a result, all key bits matched to such base bits are discarded by the Transmitter and Recipient. If b = bi, every qubit is produced and computed on another premise as the originator, resulting in the recipient receiving the crucial bit encrypted by the recipient, k = ki transmitting the concealed key (unless someone eavesdropped).

3.5. Encryption Process of Messages

The encryption bits are transformed to real bits during first stage. The bits are therefore translated into Radio signal. Finally, the communication was authenticated.

4. Results

A crucial indication for assessing key pre-distribution schemes is the likelihood whether two neighbouring nodes will create a transmission key simultaneously and the comparison done in [23]. The likelihood c of nearest neighbors is inversely proportionate to the quadratic key source, as seen in Figure 3. The network's connection likelihood grows. The network's connection can be improved by assigning the same number of keys to every station and doubling the value of produced variables. The network connection

has been enhanced as a result of our approach. Figure 4 shows how people communicate when a connection is made.

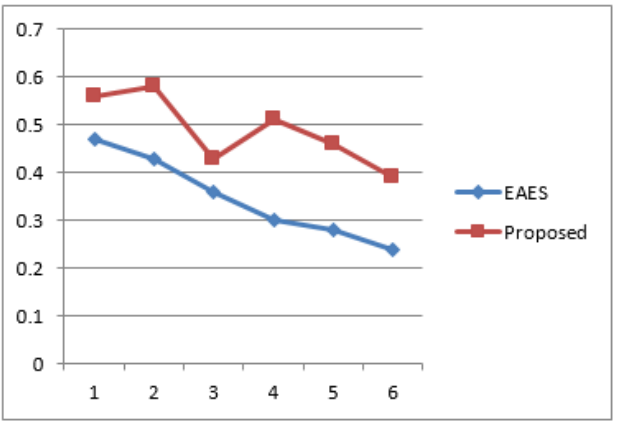


Figure 3. Quadratic key source.

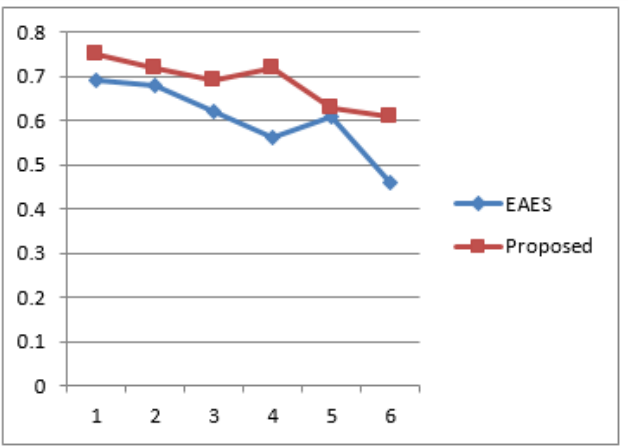


Figure 4. Shows how people communicate when connection is done.

The ciphertext from server to client transmits photonic in randomized symmetric or orthogonal patterns, that are then transformed into protected key bits.

The cryptographic key bits are received by the host, which are then decoded on identical randomized symmetric or orthogonal bases, yielding the initial message. In Figure 5, the device sends a confidential response to the

client, which is likewise encoded and decoded in the very same randomized symmetric or orthogonal bases at the receiving end, as well as the original statement obtained.

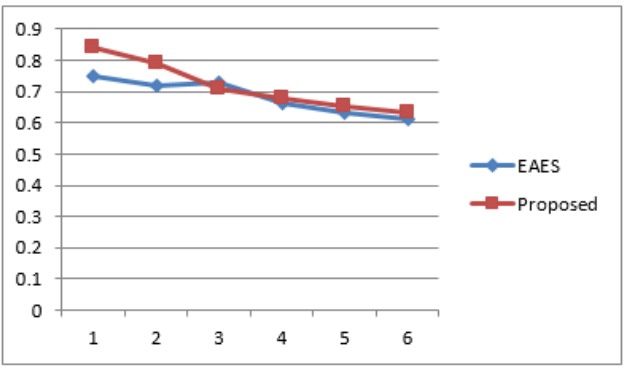


Figure 5. Device sends a confidential response to the client.

Conclusion

The QKD technique is used to increase data transmission privacy. In the randomized straight line and orthogonal bases, the QKD creates visible light for encoding and decoding.

The method enhances the safety of quantum systems beyond that of computational methods. In a wireless link, the Swift code being used to encryption and decrypting is more successful in changing bits into visible light.

Information may be transferred now that connectivity between the server and the provider and the consumer has been achieved. Furthermore, the technique may be used on any sort of connection.

References

[1] Peng Yan and Nengkun Yu. "*The QQUIC Transport Protocol: Quantum assisted UDP Internet Connections*" arXiv preprint arXiv:2006.00653 in June 2020.

[2] Zhao G. 2011. *Network Protocols and Algorithms*, (Penza: PSU) 3(1) 46-63.

[3] Kim HJ. 2012. *Int J of Distributed Sensor Networks*, (Hindawi Publishing Corp.) 8(11) pp 1-10.

[4] Kravets AG, Bui ND and Al-Ashval M. 2014. *Communications in Computer and Information Science,* (Springer IPS) 466 371-382.

[5] Bolshakova ON and Chusavitina GN. 2016. *Proc 3d Int Conf on Information Technologies In Science, Management, Social Services and Medicine,* (Tomsk: TPU) pp 122-126.

[6] Finogeev AG, Maslov VA, Finogeev AA and Bogatyrev VE. 2011. *Proc VSTU (Actual Problems of Management, Computer Science and Informatics in Technical Systems),* (Volgograd: VSTU) 3(10) 73-81.

[7] Kizim A, Matokhina A and Nesterov B. 2015. *Communications in Computer and Information Science,* (Springer IPS) 535 354-366.

[8] Finogeev AG, Dilman VB, Maslov VA and Finogeev AA. 2011. *Applied Informatics,* (Moscow: Market DS) 3(33) 83-93.

[9] Sokolov A, Tyukov A, Sadovnikova N, Zhuk S, Khrzhanovskaya O and Brebels A. 2015. *Communications in Computer and Information Science,* (Springer IPS) 535 462-473.

[10] Laszlo Gyongyosi, Laszlo Bacsardi and Sandor Imre, "A Survey on Quantum Key Distribution" in *Infocommunications Journal*, volume XI in June 2019.

[11] Alharith A Abdullah, Rifaat Z Khalaf and Hamza B Habib. "Modified BB84 Quantum Key Distribution Protocol Using Legendre Symbol"in *2nd Scientific Conference of Computer Sciences (SCCS), IEEE,* 2019.

[12] Aditya Sharma, Ila Sharma and Aaditya Jain. "A Construction of Security Enhanced and Efficient Handover AKA Protocol in 5G Communication Network" in *10th ICCCNT 2019,* IIT - Kanpur, Kanpur, India.

[13] Awais Javed M and Sohaib khan Niazi. "5G Security Artifacts (DoS/DDoS and Authentication)" in *2019 International Conference on Communication Technologies,* (ComTech 2019).

[14] Kaihong Han, Maode Ma, Xiaohong Li, Zhiyong Feng, Jianye Hao. "An Efficient Handover Authentication Mechanism for 5G Wireless Network" in *2019 IEEE Wireless Communications and Networking Conference (WCNC).*

[15] Madhusanka Liyanage, Jukka Salo, An Braeken, Tanesh Kumar, Suranga Seneviratne, Mika Ylianttila. "5G Privacy: Scenarios and Solutions" in *2018 IEEE*.

[16] Jing Yang, Xinsheng Ji, Kaizhi Huang, Yajun Chen, Xiaoming Xu, Ming Yi. "Unified and fast handover authentication based on link signatures in 5G SDN-based HetNet" in *IET Commun,* 2019, Vol 13 Iss. 2, pp 144-152.

[17] Ijaz Ahmad, Tanesh Kumar, Madhusanka Liyanage, Jude Okwuibe, Mika Ylianttila and Andrei Gurtov. "Overview of 5G Security Challenges and Solutions" in *IEEE Communications Standards Magazine,* March 2018 pp. 2471-2825.

[18] Shailesh Pramod, Bendaleand Jayashree, Rajesh Prasad. "Security Threats and Challenges in Future Mobile Wireless Networks" in *2018 IEEE Global Conference on Wireless Computing and Networking (GCWCN).*

[19] Sooyeon Shin and Taekyoung Kwon. "Two-Factor Authenticated Key Agreement Supporting Unlinkability in 5G-Integrated Wireless Sensor Networks" in *IEEE Transactions*, Volume 6, 2018.

[20] Ting Ma, Feng Hu and Maode Ma. “Fast and Efficient Physical Layer Authentication for 5G HetNet Handover” in *2017 27th International Telecommunication Networks and Applications Conference*.

[21] Rodriguez Jonathan, Radwan Ayman, Barbosa Claudia. “SECRET - Secure Network Coding for Reduced Energy Next Generation Mobile Small cells” in *IEEE A European Training Network in Wireless Communications and Networking for 5G*.

[22] Mazin Asim, Davaslioglu Kemal and Gitlin Richard D. “Secure Key Management for 5G Physical Layer Security” in *IEEE Transactions*, 2017.

[23] Giji Kiruba and Benita. "Energy capable clustering method for extend the duration of IoT based mobile wireless sensor network with remote nodes" *Energy Harvesting and Systems*, vol no 2021, pp 000010151520210006. https://doi.org/10.1515/ehs-2021-0006.

[24] Kanaga Suba Raja S, Usha Kiruthika S. “An Energy Efficient Method for Secure and Reliable Data Transmission in Wireless Body Area Networks Using RelAODV” in *International Journal of Wireless Personal Communications*, 2015.

[25] Xiaoyu Duan and Xianbin Wang. “Authentication Handover and Privacy Protection in 5G HetNets Using Software-Defined Networking” in *IEEE Communications Magazine,* 2015.

[26] Jin Cao, Maode Ma, Yulong Fu, Hui Li and Yinghui Zhang. “CPPHA: Capability-based Privacy-Protection Handover Authentication Mechanism for SDN-based 5G HetNets” in *Journal of Latex Class Files,* vol 14, no 8, august 2015.

[27] Rajakumar Arul, Gunasekaran Raja, Alaa Omran Almagrabi, Mohammed Saeed Alkatheiri, Sajjad Hussain Chauhdary and Ali Kashif Bashir. “A Quantum-Safe Key Hierarchy and Dynamic Security Association for LTE/SAE in 5G Scenario” in *IEEE transactions on industrial informatics,* vol 16, no 1, pp 681-690, January 2020.

[28] Degefa BF, Lee D, Kim J, Choi Y and Won D. “Performance and security – enhanced authentication and key agreement protocol for SAE/LTE network” in *Computer Network,* vol 94, pp 145-163, 2016.

[29] Qiu Y, Ma M and Wang X. “A proxy signature – based handover authentication scheme for LTE wireless networks” in *Network Computation Application Journal,* vol 83, pp 63-71, 2017.

[30] Rahim Robbi, Murugan S, Reham R Mostafa, Anil Kumar Dubey, Regin R, Vikram Kulkarni and Dhanalakshmi KS. “Detecting the Phishing Attack Using Collaborative Approach and Secure Login through Dynamic Virtual Passwords”. *Webology*, 17, no 2 (2020).

[31] Rajesh D, Jaya T. “Exploration on Cluster Related Energy Proficient Routing in Mobile Wireless Sensor Network”. *International Journal of Innovative Technology and Exploring Engineering (IJITEE),* vol 8, 2019.

[32] Rajesh D, Kiruba DG. “A probability based energy competent cluster based secured ch selection routing EC2SR protocol for smart dust”. *Peer-to-Peer Netw Appl,* (2021). https://doi.org/10.1007/s12083-021-01144-z.

[33] Giji Kiruba D and Rajesh D. “Analysis on clustered energy organized in mobile wireless sensor networks”. *International Journal of Engineering and Robot Technology*, vol 5, pp 37-42, 2018.

[34] Rajesh D, Jaya T. “A Mathematical Model for Energy Efficient Secured CH Clustering Protocol for Mobile Wireless Sensor Network”, *Wireless Pers Commun*, vol 112, pp 421-438, 2020.

[35] Rajesh D, Jaya T. “ECIGC-MWSN: Energy capable information gathering in clustered secured CH based routing in MWSN”, *Materials Today: Proceedings*, https://doi.org/10.1016/j.matpr.2020.09.100, 2020.

[36] Rajesh D, Jaya T. “Enhancement of network lifetime by fuzzy based secure CH clustered routing protocol for mobile wireless sensor network”, *J Ambient Intell Human Comput*, 2021. https://doi.org/10.1007/s12652-021-03170-4.

[37] Lu ZM and Mohamed H. (2021). A Complex Encryption System Design Implemented by AES. *Journal of Information Security*, 12, 177-187.

[38] Rajesh D, Firoja Banu M, Stella D, Grace Ansila P. “Ch Panel Based Routing Scheme for Mobile Wireless Sensor Network”, *International Journal of MC Square Scientific Research*, 8 (2016).

[39] Zhang C, Yao Y, Shu X, Li Z, Tang Z and Wu Q. (2020). *Data-drivenMeta-set Based Fine-Grained Visual Classification*. arXivpreprintarXiv:2008.02438.

[40] Giji Kiruba D and Rajesh D. Energy resourceful routing in mobile wirelesssensor network utilizing clustered related method, *International Journal of Engineering and Robot Technology*, 4(2),2017, 41-48.

[41] Zhang C, Yao Y, Shu X, Li Z, Tang Z and Wu Q. (2020). *Data-driven Meta-set Based Fine-Grained Visual Classification*. arXivpreprintarXiv:2008.02438.

[42] Hu C, Cheng X, Zhang F, Wu D, Liao X, Chen D. OPFKA: Secure and efficient Ordered-Physiological-Feature-based keyagreement for wireless Body Area Networks. In *Proceedings of the IEEE Annual Joint Conference: INFOCOM, IEEE Computerand Communications Societies*, Turin, Italy, 14-19 April 2013; pp. 2274-2282.

[43] Rajesh D et al. A Study of Enhancing Security of Sensitive Statistical Information Using Hybrid Paradigms, *International Journal of Engineering and Robot Technology*, 1(2), 2014, 56-61.

Chapter 3

Quantum Computing of PMS Using Machine Learning Algorithms for Revenue Management in Front Office Operations

H. M. Moyeenudin[1] and R. Anandan[2]

[1]School of Hotel & Catering Management,
Vels Institute of Science, Technology and Advanced Studies,
(VISTAS), Deemed to be University, Pallavaram, Chennai, India
[2]School of Engineering, Department of Computer Science Engineeering,
Vels Institute of Science, Technology and Advanced Studies,
(VISTAS) Deemed to be University Pallavaram, Chennai, India

Abstract

The growth of hotel industry has a significant impact on technology, especially with the Property Management System of hotels. There is a need of business management using recent technologies to improve their volume and quality of business. The contemporary Quantum Computing (QC) which has made substantial advances in a variety of fields with the aid of machine learning, including cryptography to secure communication and solving big data challenges using machine learning. Quantum Machine Learning (QML) is a new field in the quantum world that combines conventional Machine Learning (ML) with quantum information processing (QIP) to tackle and improve problems that are seen in classical machine learning (e.g., time and energy consumption). In future, quantum computing is increasingly being considered to meet these computational needs. Quantum gadgets and small scale quantum computers has to be developed. There are companies that have already deployed quantum computing which is being marketed commercially. These Quantum computers have the potential to benefit machine learning research and applications in all fields of science and engineering. The

In: Frontiers in Quantum Computing: New Research
Editor: R. Anandan
ISBN: 978-1-68507-816-4
© 2022 Nova Science Publishers, Inc.

purpose of this research is to evaluate the use of Quantum Computing in the field of front office operations of hotel industry for a better Revenue Management (RM). Secondly on knowing the compatibility with Property Management System (PMS) of a Luxury Hotel for secure and safe payments.

Keywords: quantum computing, PMS, machine learning, algorithms

1. Introduction

Luxury Hotels have various products such as rooms, buffet, spa, bar and many more, Revenue Management (RM) is a bunch of strategies used in regard with hotel business especially in front office operations that carries out to sell rooms for potential customers, as like in airline industry were the travel apps that support in booking air tickets, thus hotel industry can attract ideal customers for their rooms with the help of PMS in integration with Machine Learning (ML) [1], which is a perfect opportunity for hotel to sell their rooms at the right cost. This intricate trend in hotel industry creates the demands of PMS that is shaped by the request of client and also based on economic situations. The client conduct and socioeconomics conditions will advance the change in PMS according to the demand, thus to evaluate and utilize the sources in a profitable way through the dynamic valuing procedures of Quantum Computing (QC) hotel PMS must get integrated with QML. The implementation of QC is a future option for dynamic evaluation on hotel room pricing based on trend [2]. The tourism industry also offer facilities like aircrafts and vehicle rental to change costs on the fly by continually recalculating the interest on supply and demand. In addition the tourism board also holds properties for accommodation that has to adapt the recent technologies for a better RM. Nonetheless, in view of basic fundamentals in quantum computing, which is concentrated towards the revenue management using PMS at hotel Industry has remained generally limited to the material science of the local area, with most of work being carried out to reach the customers of different places. Dynamic valuing approach is applied by luxury hotels for pricing, is just as the initial stage of this century, as the international chain group of hotels like Hilton, Holiday Inn, Hyatt, Marriott, and other Hotels has already started executing RM programming frameworks for their properties. The technology in hotel industry uncovers that 60% of luxury hotels utilizing PMS along with revenue management software. Besides, as

per various assessments results shows there is a drift in hotel business due to lack of information in regard with technology [3]. Restaurants and Hotels have been using RM methods and techniques in past decades are not challenging, the traditional revenue management technique have become out dated with the legacy of manual methods. As the trend advanced with android and Internet of Things (IoT). The hotel business shift from basic inheritance frameworks to more intelligent information driven arrangements through Quantum Computing. In this article, the novel approaches are examined for a better yield through hotel front office that proves how information examination considers building exact and powerful unique valuing applications through ML in addition with QC. The hotel PMS are regularly confounded and actually inadequate. For example, the hotel front office managers should physically change costs, restricting the spryness of the unique estimating approach [4]. The issue stretches out to different activities like during the guest stay at hotel at times rooms might be allocated exclusively by the purpose of stay along with the number of days which is not precise as it tend to change for various reasons. The hotels rooms most often reserved through Online Travel Agent (OTA) websites, thus the authentication of guest arrival has also become a challenge as it varies according to time and also with guest interest, there are times were hotel use to have continuous reservation cancellations [5]. Therefore to avoid ambiguity and to determine the exact occupancy in order to manage the hotel capacity the machine learning algorithm is used. The convenience of hotel predictions generally goes with the previous year response rate that might be identified through straightforward techniques. Because of old programming execution on PMS, thus information handling is deferred in recent years as it based on IoT aided with OTA's, Hotel apps and websites. The data on inventories is every now and again conflicting with variations thus application of QC will provide an accurate results in addition with ML [6].

2. Literature Review

The Luxury Hotel work is regularly coordinated with each department accordingly at times the information's are shared for better performances using IoT, further the hotel property is managed with a PMS in coordination with OTA to attain administrative strategies [7]. Thus at times there are issues and these issues lead to unnecessary estimates and consumption of time, which brings about low inhabitance and critical income misfortunes for hoteliers.

The modernization of income in hotel frameworks can altogether change the circumstance, by connecting new modules with a PMS, hotels can forestall overbooking and selling beneath costs. The yield management standards continue as before paying little mind to basic programming. In any case, an unmistakable quality shift began in the hotel business as Artificial Intelligence (AI) and information science-based on strategies were presented for yield management with programming [8]. AI involves building and preparing measurable models utilizing information contributions to arrange input data and yield output with nonstop qualities [9]. There are two regularly utilized sorts of ML execution: one is supervised ML another one is unsupervised ML. The previous requires a noteworthy dataset with named yield approvals to put together expectations with respect to the new information. The last option that is unsupervised ML utilizes unlabeled information to track down association between various properties [10]. As far as yield management with the supervised learning methods which proceeds into consideration anticipating best-fit estimating, consequently characterizing yield strategies and gauging request [11]. Unsupervised ML can be applied to various exploration assignments, for example, making profoundly determined division that records for suppressed associations like client socioeconomics inclinations.

3. Quantum Algorithm for Property Management System

To solve the issues raised in recent time quantum chips are the best choice which has an unbelievable amount of computational command and control. Thus to performing task by quantum computers using machine learning is more beneficial with integration of Application Specific Integrated Circuits (ASIC). The most significant application to perform data analytics for knowing the best option is machine learning and artificial intelligence in these present days which is possible by integrating the computer with Field Programmable Gate Arrays (FPGA) that are intended to facilitate ML [12]. Making a PMS based on the requirement of Luxury Hotel according to the size and kind of the property is essential, since it is costly not all hotels can afford to implement a PMS in their property. However to manage hotel in an efficient way with recent demands, a PMS has to be designed to meet the future demands. Many hotels follow a semi-automated PMS which contains a structure of various modules that are intended to help the hotel associates to work with the capacities. PMS programming’s are distributed to front office operations as per prerequisite and need of their association. Micros, Fidelio,

Amadeus, and IDS Fortune are some significant organizations as PMS vendors. The PMS modules of front office has different segments in order to carry out the transactions and maximize the revenue of the hotel, whereas the revenue management software is mostly separated. Thus the rates of rooms has to be manually updated for most of the PMS although there are many latest advancements have come with cloud based PMS (CPMS) which has an ability to connect with IoT based sensors for Electronic Door Locking System, (EDLS) in order to provide a secure and safe environment for the guest of luxury hotel were a code for room keys are generated from the PMS which is unique for each room and for every check in, this gives an ultimate security for hotel and guest [13].

Energy Management Systems (EMS) is a typical function which shut down the operation of equipment's those are not in use for a period of time and automatically manages room temperature, air movements in public spaces and guest rooms in order to control the whole division on minimizing the power consumption based on the occupancy of the hotel [7]. Quantum computing makes PMS to connect with other autonomous framework like retail location, EMS, managing guaranteed and non-guaranteed reservations. The EMS also gives suggestions and prioritizes which room has to sell to save energy For example, a hotel has 400 rooms and 70 percent of occupancy is forecasted where rooms are assigned in lower floors and the top floor is not occupied then the power consumption in the top floor is fully minimized it turns out to be exceptionally fruitful when EMS is a mix of equipment and programming that endeavors to cut down the energy utilization in hotel and assists with further developing support tasks. The fundamental capacity of EMS is it serves to consequently close down all the activity of the hardware's after a timeframe in case it isn't in gear. It is customized to limit the use of power as per the necessity and use of the property. In this manner EMS gives suggestion and arranges which room should be offered first and foremost to save energy which is possible on implementing machine learning.

The EDLS helps the hotel to control access the guest rooms with a proper key card coded for the specific room only the individual carries that key card can enter that room. The coding made in key are unique and only function till the guest stay in that room, The applications of PMS modules deals with the current status of rooms and outfit the subtleties on that; it permits the staff to oversee hotel and floor plan with a whole design of a layout, likewise it shows the rooms under repair, rooms in long haul stay overs, add data on rates, doling out the room during enlistment, and help front office individual to connect with visitor administrations. A room module can likewise be utilized to direct

instant entry to room accessibility information during specific states of enlistment process [14]. This product likewise helps front office staff during registration and reservations. In Hotel, now and then rooms are blocked for renovation, fix and for upkeep work or cleaning method, the quantity of rooms accessible in the booking show is naturally diminished to the real accessible rooms, this module likewise helps with controlling future room availability and guarantee that all guest have rooms when they register during check in. The Guest account module keeps the information's of hotel guest payments and the transactions made during their visit, which works on the regularly through the visitor records at hotels [15]. The visitor reservation module screens foreordained visitor credit limit and brings out flexibility through split, different folio arrangement and avoid the need of a manual folio, vouchers, reports and registers. At the hour of checkout, recently approved extraordinary record adjusts are consequently moved to a proper document for future reference. At the point when the hotels, F&B Outlets, business focus on providing the facilities, some of the additional services like fitness and spa, which are associated with the front office remote retail location modules, which can impart visitor charges to QPMS with the assistance of Internet.

3.1. Central Reservation System (CRS)

Central Reservation System (CRS) are uncommonly intended for the guest who plan for travel and to visit various places, thus tourism boards are focused to empower their proficiency and efficiency, this framework gives fundamental information concerning the Hotels, Transport with journeys timetables and seats could be managed processed using this tool. Already when there was no proper conveyance framework was observed for all type of tourist, the information and data about rooms, flight, rent a vehicle organizations would require a long time to engrave as handouts and leaflets, thus now digital marketing in hotel industry through recent tools helped in information exchange strategy between airline companies, travel agents and explorer's. Bringing about late return of data from the travel planners to the traveler which took long time, using GDS and Online Travel Agents (OTA's) which provides details of visitor as it is an independent tool thus implementation of CRS will maximize the revenue through the automation in reservation further it allows the guest to manage their reservation using this facility that will simultaneously increase the loyalty towards the brand and satisfaction on the services provided, an adverse consequence on their

organization quality. The framework's subsequent capacity is to hold items or administration that makes an enrolled traveler or client name during their visit [16]. A CRS is either integrated with PMS or can be an independent software tool. It's firmly identified with the front office operations, as it passes all the data about recent reservations, and acts as a revenue management source with its modules in PMS. The PMS and all appropriation interfaces gathers data about the room accessibility and rates from the CRS and offers it across numerous channels, like GDSs, OTAs, wholesalers, search destinations, and the lodging's own site through various sorts of APIs. As rooms get booked, it synchronizes the stock data all through every one of the internet based stages to forestall overbooking and furthermore keeps up with rate equality [17]. A channel supervisor can likewise come as a piece tool along with PMS, as a feature of the CRS if the last option is planned independently, or as a free programming to be incorporated with the current lodging programming.

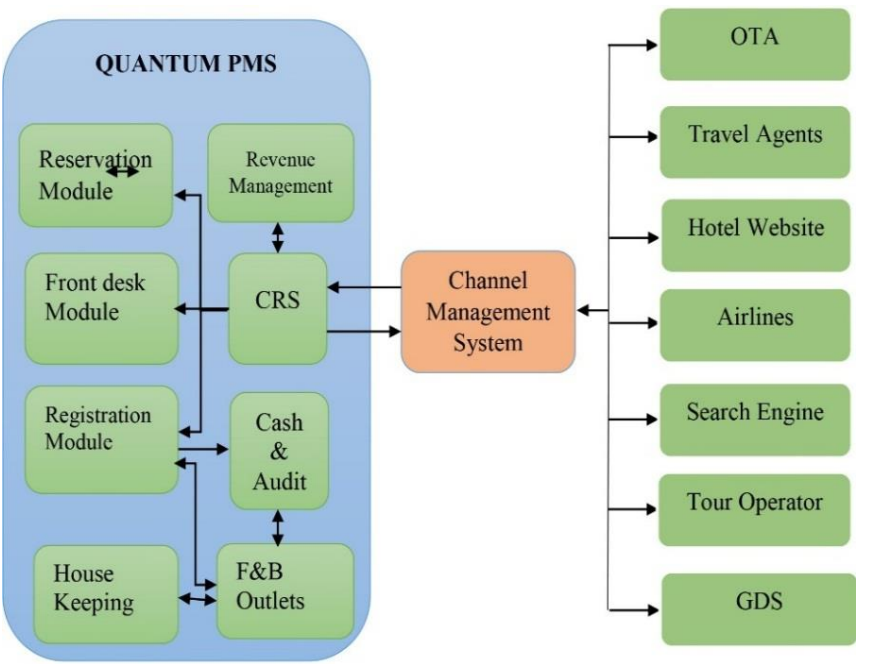


Figure 1. Quantum PMS.

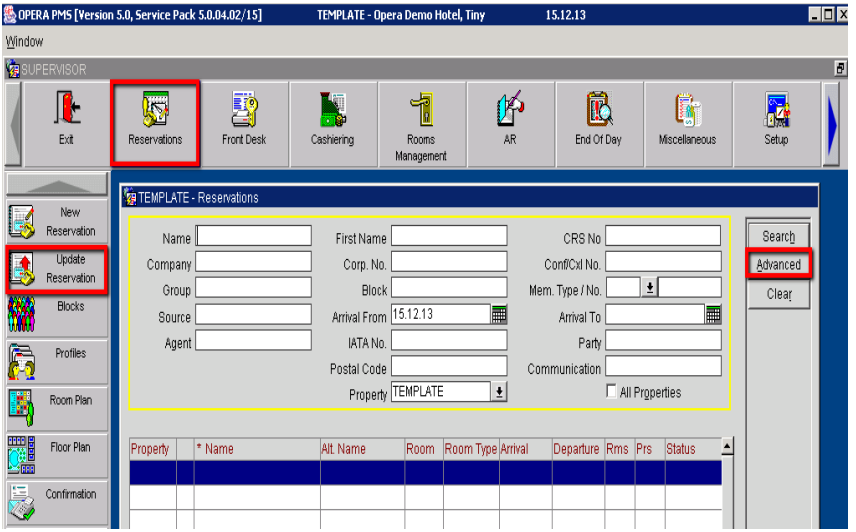


Figure 2. Front desk PMS main screen.

3.2. Applications of PMS

The hotel Property Management Systems have made considerable progress to become one of the most fundamental innovation apparatuses for hoteliers in recent years. Basically a PMS is seen as the spine for the properties like resorts, luxury hotels, some of the PMS make it more convenient for organizing the sections of the front office like reservation, registration, information and cash, front office executive's deals with arranging the guest stay in hotel through taking reservations and making registration during the arrival of guest at hotel [18]. Thus, PMS permits the hotel administration to know the occupancy and receiving the guest payments towards the guest stay at hotel, further it is equipped to confirm the guest reservation through online, To do that there are applications or Modules of PMS at front desk like Reservation Module, Registration Module, Cash Module. During the reservation guest enters the details of them like, name, address, phone number and their duration of stay, mode of payment, etc. In order to get a satisfied guest the preferences has to be recognized, which is not possible with a classical PMS [19]. There are some recent PMS that has been equipped with EMS, EDLS, having option for blocking the room that is under repair or out of order, creating room key code during check in, making miscellaneous transactions on requirement, in addition there are options for making

transactions through Food & Beverage (F&B) outlets, like Banquets, Restaurants, Coffee shop, Room service, etc. PMS provides the access to do all these transactions of hotel, but in general the quality is obtained by the satisfaction of the guest who visits the hotel premises. Therefore these application are intended for the ease of hotel operations, whereas the guest expectations are full filled on choosing the right place for their stay, understanding the guest preferences and providing a right room to them is possible now by QML which has to be integrated by the PMS to make it as QPMS, in future hotels can obtain a maximum satisfaction from the guest in terms of quality and meeting the demand of their hotel guest, Thus, the rates of rooms must be competing with hotels which has similar facilities and services to make the guest to feel that it is value for money. Machine Learning with Quantum Computing will provide an output which is going to be more accurate and supportive to manage the hotel business.

Implementation of QML in PMS will provide a QPMS, This portrayal would be made in a period significantly less than the quantity of items used though an algorithm like Grover's or Shor's with a values like N somewhere close to O (1) to O (logN). In the traditional case, this information would just be put away in Random Access Memory (RAM) and called when required. Notwithstanding, while RAM is a vital component of traditional figuring, as of now no hearty viable RAM identical that produces the required quantum superposition state for a quantum PC. It has been recommended that a quantum rendition of irregular access of RAM, or QRAM, could create this information in O (log N) time, albeit this has not been essentially illustrated. To accomplish this, a traditional information stockpiling unit would be enhanced with quantum validation among the available data on memory. A traditional simple to this design, called a Content Addressable Memory (CAM), this will exists and takes care of this pursuit issue in O (log N) time. In any case, with CAM and with QRAM, getting the information into the gadget in any case actually takes O (N) time, so either approach might be valuable when various studies are performed on similar data that is, the utility of CAM and QRAM, assuming it tends to be fabricated, fills in direct extent to the occasions the information can be reused for better performance.

3.3. Reservation Module

The main advantage of a PMS is that it assists hotels business by its features as it consistently overseeing reservations of hotel guest. An ideal PMS ought

to permit a hotel to concentrate all requirements of the visitors not only the reservation, but making it guaranteed reservation through online payments, updating the guest with automated emails on reservation, further in connecting with websites and OTA's for reservation. At this point when every one of the reservations are reflected in one spot, it becomes more straightforward for the staff of front office in hotel to know the exact occupancy during check in. the social separating turning into the standard, computerized or contactless registration is an inclination for some visitors these days. The contactless innovation permits visitors to show up, and afterward registration utilizing the portable application on their telephone, or by utilizing a self-administration registration machine. The PMS can synchronize consequently, keeping registration and inhabitance data refreshed [20]. It is basic requirement for hotels to gather and arrange visitor information to stay in contact with current and past clients during and after guest visit at hotel. A CRM incorporated with PMS can help the front office operations through group reservation to gather all visitor data from these sources, the guest information have been stored in a data set with an open configuration. Progressed QPMS likewise empowers hotels to oversee visitor devotion programs successfully. Hotels in today trend need various innovative solutions to deal with their property effectively. In any case, the vital challenge for hotel is that most advanced PMS are not providing solutions managing reservations, and it becomes hard for hotels to get an unmistakable image of how everything can cooperate. A QML with PMS can assists hotels by having synchronized information fundamental for informed independent direction. Thus it is significant for a hotel PMS to have open API or choice to coordinate to other reservation frameworks so hotels can take advantage of it.

Most hotel PMSs work with different channel administrators or Channel Management System (CMS), as rate fixed by the revenue management system of hotel seems to be competing to make it consistent for more reservation in to hotel, furthermore these rates are also available for accessibility in OTAs platform and get reservation of hotel rooms reflect in PMS. Some PMS have options like channel manager which permits luxury hotels to offer better online reservation experience through a tremendous volume of occupancy in hotel through the web deals channels. since retraction designs change over the long haul and in light of the fact that the framework was needed to adapt consistently, a weighting component was made to ascribe higher significance to ongoing confirmed reservations and to obtain a rate penetrating learning methods as a visual cue weighting dependent on the predictions made previously accurately and mistakenly [21]. A booking can be anticipated as

cancelled at times, however as not cancelled on the following day. Estimating the accuracy of past forecasts on unsteady perceptions required the improvement of another action, Minimum Frequency. To conquer this weakness and to empower the framework to consistently gain from new information, the framework was intended to fuse the Quantum Algorithm. Rather than allowing the model to give decreased performance, a new model could be generated through QML. This model could be based on consistently delivering and exhibiting the contrasted presentation of the current model. The model with predominant outcomes will be chosen.

3.4. Revenue Management System (RMS)

The front office PMS could additionally assists hotels with overseeing income successfully. It assists hoteliers with surveying the presentation of the property, utilizing key execution points like Average Daily Rate (ADR), Revenue per available room (REVPAR) Average Room Rate (ARR) which could be obtained by the room occupied in a hotel, further the room rates are keep changing on knowing the competition in the industry, whereas the rates are keep fluctuating based on demand at OTA websites. Thus, now there is a great demand for Quantum computing and machine learning which assist in fixing the room rates according to the demand as well as for the benefit of the organization. A decent PMS additionally offers rate plan of the hotel where lodgings can rapidly make room type and rate, which helps hoteliers to effectively react to changes sought after. These adjustment of awareness might happen because of irregularity, locality events, and different variables raised during the time of guest check in.

3.4.1. Capacity Management

This type of approach is to assign the rooms based on demand and to control for optimum profit, by reserving the excess rooms than the available number of rooms in hotel for e.g., a hotel has 200 rooms have taken reservations for 250 in order to adjust the rooms that are going to be cancelled, thus in hotel PMS there are options to know the number of rooms reserved, whereas the cancellations are not predicted. A machine learning algorithm can predict the occupancy on previous and present data but with the situations like flight delay or cancelled status could be identified by the algorithms if that is connected with the airline data base, further it will helps in assigning the rooms to other

guest who wish to reserve a room in that hotel. This way of approach will assist in avoiding potential loss to the hotel due to the unoccupied rooms.

3.4.2. Discount Allocation Based on Demand

A front-office module permits discounts on special occasions or for Travel agents and Tour operators, in addition the discounts are given for Airline companies, corporate companies and to V.I.Ps. But there are seasons when the hotel occupancy reach to 20 to 25% to increase the occupancy the rates has to be competing with other players. Some PMS are semi-automated even in fully automated there are no provisions for rates revisions automatically as it has to be decided by management of the particular hotel. PMS offer facilities like reconciliation into registration desk or permit check in by means of QR code. There are module associates with PMS are reservation, housekeeping, spa, Health and fitness, for making discounts the PMS must have a revenue management system which is aided with Machine learning.

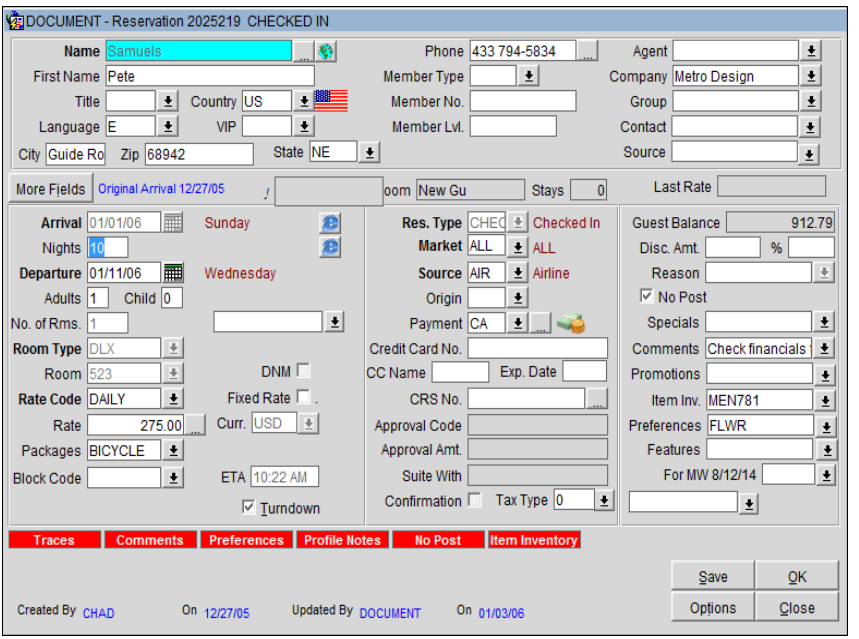


Figure 3. Reservation module of PMS.

3.4.3. Duration Control

Reservation of room in hotels and resorts based on situations and demand allocating rooms for the guest who wish to stay in hotel for longer duration has been given priority, mostly the discounts are given on rooms because it is considered as a most perishable product [22]. Thus the rooms are offered in discounted rates which is less than the rack rate, also selling a product which is highly perishable in a discounted rate is a wise decision. The aim of duration control is to provide the space for the guest who wish to stay for longer time, were you can avoid taking reservation for one day or half day as the hotel room will going to be occupied on the next day will be for longer period. There are situations were hotel could not able to check in a group because there is shortage of one or two rooms this could be avoided without taking reservation or controlling by blocking the rooms for group check in, at some circumstances hotel provide rooms on discounts if it is a four or five days stay whereas for one day guest must go with a rack rate, as one day stay of guest may become detrimental for the overall revenue of the hotel. Thus, hotel PMS goes with automated reservations have the rare options on blocking the rooms for group, which could be possible by a revenue management software integrated with PMS and creating an opportunity for QPMS for making decisions during this crucial periods.

3.5. Global Distribution System (GDS)

There are networks like Global Distribution System (GDS) which is computerized network that sends availability and charges of the hotel establishments to the development trained professional of tourism and travel agents. The use of GDS on the fields of tourism can extend the availability of accommodation, return and swiftness of correspondence among the associations in giving the organizations to the potential guest is mandatory. GDS is mostly consolidated among significant Airlines Company, vehicle rental organizations and Hotel gatherings. The aftereffect of present day current data innovation, GDS are particularly vital as they empower the guest to get the information from any spot on the planet and proposition them the opportunity to reserve a spot in a superior manner that steadily take an interest in complete hotel reservations through cooperative technique and pushing the hotel in the movement market [23].

3.6. Online Travel Agency (OTA)

OTA is an online commercial websites that permits travelers to plan and book their travel requirements like transport by train, airline or cruise line, accommodation in hotels, motels or resorts, in addition it helps in managing their trips according to their preferences on selecting their choices, and that's just the beginning, honestly with tourists. Consistently, there are many explorers throughout from globe use OTAs to design their trip for recreation and business travel. OTAs give admittance towards their visit with the possible highlights of a location and at volumes that would be hard to access through any other advertising medium. Moreover, OTAs give market bits of knowledge along with the devices to focusing on explorers expectation and in getting and handling appointments, speaking with users and overseeing surveys with reviews, which is based on hotel facilities and sizes including bed and morning meals, comfort, luxury, hygiene and customer satisfaction, further more options are available for motels and resorts for those properties who have signed agreements to list their properties on an OTA Platform [24]. There are no forthright charges to list rooms in the commercial websites, thus OTAs gets their service charge in percentage only when rooms are reserved and in order it provides a platform to display their property images like rooms, restaurants, spa, fitness centers and swimming pool through OTA with a level of quality to complete worthy reservation. The OTA gives freedoms to feature rooms and properties through convenience depictions, photographs of landmarks based on interests through the filters and that's only the tip of the iceberg. Postings are shown to explorers looking for spots to remain in their objective. Assuming that they use of channels to limit their pursuit for example just searching for swimming pools, free Wi-Fi, Bar. OTA will adjust travel rules with the detail given by the luxury hotel by showing the properties that match best with expectation. The OTA gives a safe booking stage and is the central matter of contact for booking corrections and scratch-offs. At no expense, OTAs give an assortment of tools from continuous admittance to advertise information to gadgets for overseeing audits to assist their accomplices with being effective. It's getting harder to track down an example of overcoming adversity of how travel organization like Hotel X made innovation improvement in their property further Y has expanded deals and achieved some percentage of profit without recognizing the words ML of Artificial Intelligence (AI). Once in a while utilized conversely, these two ideas really have various implications. AI is an immense space of software engineering that concentrates on the best way to train the data which is

possible by algorithms that has the ability to think and behave like a human. The subset of AI is ML, which is significant in building models, yet not alone. More or less, ML is tied in with building models that foresee the outcome with the high precision based on the information. Utilizing measurable strategies, it empowers machines to work on their exactness as Big Data is taken care in the framework [25].

The Big Data is assorted to make it enhanced, these data has to be trained and tested, the better the machine learning models can track down designs and the more exact the outcome will come. For instance, here certain ways how and which information can be caught by tourism industry. The datasets of good quality are as a rule in exceptionally appeal and the organizations once in a while in a real sense need to chase after the fair datasets. Normally information has more data that is expected to construct the model, so it is important to choose the significant elements. During this interaction, either the examiner or demonstrating instrument chooses or disposes of the characteristics relying upon how valuable they are for investigation.

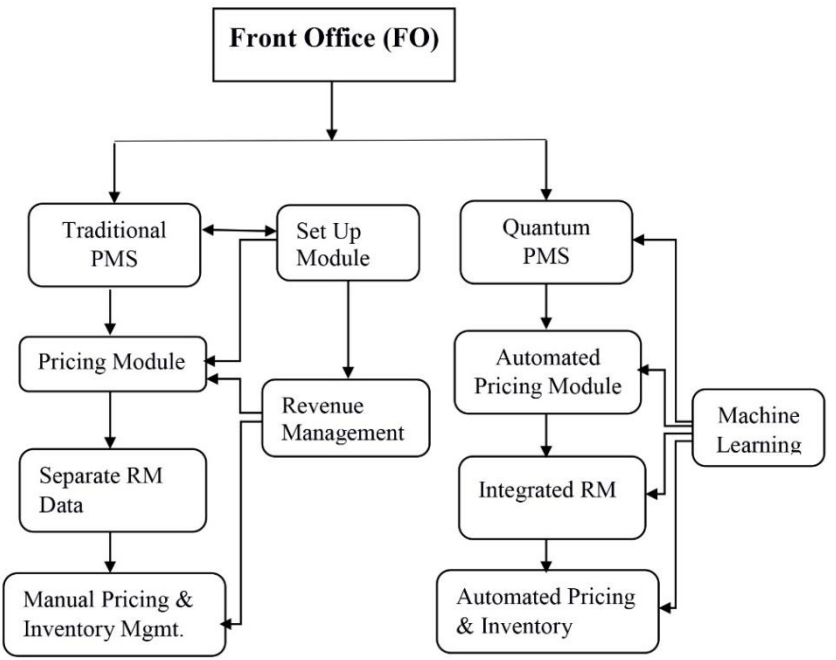


Figure 4. Features of PMS.

3.7. Quantum Property Management System with OTA

The channel Management System of QPMS could automatically connect with OTAs websites for reservation of hotel rooms, the dynamic pricing strategy from inbuilt Revenue Management System (RMS) can able to react according to the supply and demand, this software is going to be customizable with its user friendly features. A large number of tourist attractions make the calculations to work more slow, so generally the course of information arrangement and having files like .xlsx with various parameters of hotels facilities and the services provided, similarly with .csv records in the end takes additional time than the entire course of preparing results [26]. Thus ML that examines the information searches for examples or patterns and afterward tracks down the ideal boundaries for making the model. It's a significant test to pick the best algorithm to settle a particular assignment as every calculation that can produce an alternate outcome and some of them create more than one sort of result. Machine learning model can beat classical style unbending business knowledge where business rules can't catch the secret details. OTA's are effectively carrying out AI and ML to examine somewhere down in the accessible information and enhance the stream on their sites and applications, and convey genuinely prevalent encounters according to hotels. Thus a PMS which has the facility to connect with OTA's could be a better option for luxury hotels.

3.8. Grover's Algorithm

The algorithm that works by building blocks is Quantum Fourier transform (QFT) algorithm is termed as Grover's algorithm. The QFT triggers numerous quantum calculations, one more class of calculations exploit an alternate strategy, called the Quantum Random Walk (QRW). This technique is comparable to old style arbitrary walk strategies, which probabilistically recreate progress in navigating some territory. Thus the algorithm by Grover resolves the particular issue of finding the extraordinary contributions to a given capacity that will yield a specific output from the available data. Traditionally, this is an essential Polynomial Time Solution (PTS) [27]. During the non-availability of data and about the idea of the capacity, the quickest known traditional algorithm to solve this issue is thorough investigation of all potential contributions to find the appropriate response when a cycle that makes in strides, where n is the quantity of pieces needed to

address the data. Grover's calculation tackles this issue in Image steps. While this is just a polynomial speedup over the best traditional approach, it could regardless be huge practically speaking. The adequate data used to think twice about cryptographic tasks. Besides, this is the ideal Quantum Algorithm (QA) for this issue, showing that any QA should find basically Image ways to tackle this issue in the model based on black box. The issue with the traditional extensive search approach is that the classical testing of every conceivable answer is a visually impaired supposition and check each question gives no data about the appropriate response until it is really found. To get around this issue, Grover's calculation continues by emphasizing a bunch of two procedure on the data. The first adequately provides the state relating to the right reply by changing the indication of its coefficient, another one is Grover Diffusion Operator, which could expand the size of this coefficient somewhat, use of these applications could expands the possibility towards the right answer, which will be perused out upon estimation. This strategy to increase the generosity of the state containing the right answer is an illustration of an overall algorithmic methodology is for the relationship valuation between the variables, which is valuable in various QA. The arrangement of relationship intensification tasks can be considered by QRW; be that as it may, in reverse, from a dispersed state that is closely resembling by all potential start and endpoints from a given stage. A traditional approach can investigate a region relative to the square base of the quantity of steps; the QRW can investigate a region corresponding to the quantity of steps [28]. Thus, the QA gives quadratic speedup in performing tasks. This strategy is exceptionally adaptable and has prompted various QA giving polynomial speedups to explicit computational assignments. For instance, there is a QRW based algorithm for tackling the essential issue of deciding if the competitor in hotel industry taking the principal action that has a triumphant technique during low demand.

The traditional algorithm includes a theatrical review of the potential changes and results could be an option using decision tree, while the algorithms used with quantum computing gives the quadratic speedup in comparison with the results portrayed previously, Thus, the usage of quantum computing in the field of customer care especially with hotel industry is essential [29], which gives a quadratic speedup to assessing the details and reviews posted by guest who stayed in hotels and their feedback on foods become more enlightening for future guest who wish to visit the hotel. While Grover's Algorithm is frequently alluded to as quantum "search," To play out a genuine quantum search, the arrangement of information to be looked

through initially and addressed as a superposition of quantum states and for a quantum computing to give rapid results.

Conclusion

This investigation of this study is centered on the implementation of QML in PMS which is intended for luxury hotels. The execution of QPMS with the ML in functional method and departmental performance of task at hotel industry is to obtain a nominal profit with customer satisfaction. This logical examination presents that the hotel industry is facing a huge competition especially during this pandemic situation, machine learning framework through the coordination of PMS will provide a better revenue with the precise measures, QPMS additionally upholds in fostering the hotel experience through different application in integration with revenue management. The revenue of the hotel need to be improved with an inhabitance by applying techniques is pivotal, while carrying out these tasks for the fulfillment of visitor satisfaction who stay in hotel is also monitored with their feedbacks based on the usage of rooms. The PMS interface through machine learning is the only way to know the guest experience from the big data and in order to compete in future with other hotels, the quantum machine learning approach is having the ability to speed up the functions in selecting the channels accordingly with requirements of hotels. The QPMS dealers should recognize the need on emerging and improving their PMS which has the capacity to exchange the information by the available data in a secure and safest way and that software will adapt the changes and amendments in all departments of hotel.

References

[1] António, N., de Almeida, A. & Nunes, L. (2019). Big data in hotel revenue management: exploring cancellation drivers to gain insights into booking cancellation behavior. *Cornell Hospitality Quarterly*. N/A, 1-22, https://dx.doi.org/10.1177/1938965519851466.

[2] Wang, Y., Xi, J.S. and Cheng, T. (2021) The Overview of Database Security Threats' Solutions: Traditional and Machine Learning. *Journal of Information Security*, 12, 34-55. https://doi.org/10.4236/jis.2021.121002.

[3] Maria-Cristina ENACHE, (2019) Machine Learning in Tourism Revenue Management, Annals of "Dunarea de Jos" University of Galati Fascicle I. *Economics and Applied Informatics*, pp 132 – 136.

[4] Nuno Antonio1, Ana de Almeida2 and Luis Nunes, An Automated Machine Learning Based Decision Support System to Predict Hotel Booking Cancellations, *Data Science Journal*, 18: 32, pp. 1–20. DOI: https://doi.org/10.5334/dsj-2019-032.

[5] N. Antonio, A. de Almeida and L. Nunes, (2017) "Predicting Hotel Bookings Cancellation with a Machine Learning Classification Model," *2017 16th IEEE International Conference on Machine Learning and Applications (ICMLA)*, pp. 1049-1054, doi: 10.1109/ICMLA.2017.00-11.

[6] Adil, Mohd, Mohd F. Ansari, Ahmad Alahmadi, Jei-Zheng Wu, and Ripon K. Chakrabortty (2021). "Solving the Problem of Class Imbalance in the Prediction of Hotel Cancelations: A Hybridized Machine Learning Approach" *Processes* 9, no. 10: 1713. https://doi.org/10.3390/pr9101713.

[7] H.M. Moyeenudin, G. Bindu, Anandan R, OTA and IoT Influence the Room Occupancy of a Hotel, Data Management, Analytics and Innovation, *Lecture Notes on Data Engineering and Communications Technologies*, 70, Volume 1, 2021, pp 265-274. DOI: 10.1007/978-981-16-2934-1_17.

[8] Thomas Millauer* & Matthijs Vellekoop, (2019), *Artificial intelligence in today's hotel revenue management: opportunities and risks*, *Research in Hospitality Management*, 9(2): 121–124 https://doi.org/10.1080/22243534.2019.1689702.

[9] ManuSharma, SunilLuthra, Sudhanshu Joshi, Anil Kumar, (2021), Implementing challenges of artificial intelligence: Evidence from public manufacturing sector of an emerging economy, *Government Information Quarterly*, https://doi.org/10.1016/j.giq.2021.101624.

[10] Manuel J. Sánchez-Franco, Antonio Navarro-García, Francisco Javier Rondán-Cataluña, (2019), A naive Bayes strategy for classifying customer satisfaction: A study based on online reviews of hospitality services, *Journal of Business Research*, Volume 101, Pages 499-506.

[11] González-Serrano L., Talón-Ballestero P. (2020) Revenue Management and E-Tourism: The Past, Present and Future. In: Xiang Z., Fuchs M., Gretzel U., Höpken W. (eds) *Handbook of e-Tourism*. Springer, Cham. https://doi.org/10.1007/978-3-030-05324-6_76-1.

[12] Zhang, H., Gu, M., Jiang, X.D. et al. (2021), An optical neural chip for implementing complex-valued neural network. *Nat Commun*, **12,** 457. https://doi.org/10.1038/s41467-020-20719-7.

[13] Gurel Cetin, Tevfik Demirciftci, Anil Bilgihan, (2021) Meeting revenue management challenges: Knowledge, skills and abilities, *International Journal of Hospitality Management*, Volume 57, August 2016, Pages 132-142.

[14] Moyeenudin, H. M., Anandan, R., & Parvez, S. J. (2020). Exploration on Revenue Using Pioneering Technology in Infrastructure Facilities of Luxury Hotels. In *Intelligent Computing and Innovation on Data Science*, (pp. 389-394). Springer, Singapore.

[15] Dyshkantiuk, Oksana and Salamatina, Svitlana and Polishchuk, Larysa and Komarnytskyi, Igor and Tserklevych, Viktoriia and Nedobiichuk, Tetiana, Modern

Hotel Business Management Tools (2020). *International Journal of Advanced Research in Engineering and Technology*, 11(6), 2020, pp. 226-238, Available at SSRN: https://ssrn.com/abstract=3645056.

[16] Vinod, B. Hotel retailing with attribute-based room pricing and inventory control. *J Revenue Pricing Manag*, 18, 429–433 (2019). https://doi.org/10.1057/s41272-019-00207-4.

[17] Goecke R. (2020) The Evolution of Online Booking Systems. In: Xiang Z., Fuchs M., Gretzel U., Höpken W. (eds) *Handbook of e-Tourism*. Springer, Cham. https://doi.org/10.1007/978-3-030-05324-6_27-1.

[18] Adeyinka-Ojo, S., Lee, S., Abdullah, S.K. and Teo, J. (2020), "Hospitality and tourism education in an emerging digital economy", *Worldwide Hospitality and Tourism Themes*, Vol. 12 No. 2, pp. 113-125. https://doi.org/10.1108/WHATT-12-2019-0075.

[19] Fuchs M., Höpken W. (2020) E-Tourism Curriculum. In: Xiang Z., Fuchs M., Gretzel U., Höpken W. (eds) *Handbook of e-Tourism*. Springer, Cham. https://doi.org/10.1007/978-3-030-05324-6_107-1.

[20] Nuno Antonio, Ana de Almeida, Luis Nunes, (2019)Hotel booking demand datasets, *Data in Brief*, Volume 22, 2019, Pages 41-49, ISSN 2352-3409, https://doi.org/10.1016/j.dib.2018.11.126.

[21] Woerner, S., Egger, D.J. Quantum risk analysis. *npj Quantum Inf*, 5, 15 (2019). https://doi.org/10.1038/s41534-019-0130-6.

[22] Zhang, Y., Mesaros, A., Fujita, K. et al. (2019). Machine learning in electronic-quantum-matter imaging experiments. *Nature*, 570, 484–490. https://doi.org/10.1038/s41586-019-1319-8.

[23] Ampountolas, A. and Legg, M.P. (2021), "A segmented machine learning modeling approach of social media for predicting occupancy", *International Journal of Contemporary Hospitality Management*, Vol. 33 No. 6, pp. 2001-2021. https://doi.org/10.1108/IJCHM-06-2020-0611.

[24] Dongmei Duan (2020) *Journal of Physics: Conference Series*, Volume 1544, 2020 5th International Conference on Intelligent Computing and Signal Processing (ICSP), Suzhou, China, 012067.

[25] Milgrom, P. & Tadelis, S. (2019). 23. How Artificial Intelligence and Machine Learning Can Impact Market Design. In A. Agrawal, J. Gans & A. Goldfarb (Ed.), *The Economics of Artificial Intelligence: An Agenda* (pp. 567-586). Chicago: University of Chicago Press. https://doi.org/10.7208/9780226613475-025.

[26] Sathish Babu B., Bhargavi, K., & Subramanya, K. N. (2021). Optimal Parameter Prediction for Secure Quantum Key Distribution Using Quantum Machine Learning Models. In I. Management Association (Ed.), *Research Anthology on Advancements in Quantum Technology*, (pp. 355-376). IGI Global. http://doi:10.4018/978-1-7998-8593-1.ch016.

[27] Sawerwain M., Wróblewski M. (2019) Application of Quantum k-NN and Grover's Algorithms for Recommendation Big-Data System. In: Borzemski L., Świątek J., Wilimowska Z. (eds) Information Systems Architecture and Technology: *Proceedings of 39th International Conference on Information Systems Architecture and Technology – ISAT* 2018. ISAT 2018. *Advances in Intelligent Systems and*

Computing, vol 852. Springer, Cham. https://doi.org/10.1007/978-3-319-99981-4_22.

[28] Sven Groppe, Jinghua Groppe, (2021). Optimizing Transaction Schedules on Universal Quantum Computers via Code Generation for Grover's Search Algorithm, *25th International Database Engineering & Applications Symposium*. Pages: 149–156, https://doi.org/10.1145/3472163.3472164.

[29] Amellal H., Meslouhi A., Allati A.E. (2019) Processing Unstructured Databases Using a Quantum Approach. In: Ben Ahmed M., Boudhir A., Younes A. (eds) Innovations in Smart Cities Applications Edition 2. SCA 2018. *Lecture Notes in Intelligent Transportation and Infrastructure*. Springer, Cham. https://doi.org/10.1007/978-3-030-11196-0_25.

[30] Vinod B. (2021) Artificial Intelligence and Emerging Technologies in Travel. In: The Evolution of Yield Management in the Airline Industry. *Management for Professionals*. Springer, Cham. https://doi.org/10.1007/978-3-030-70424-7_11.

[31] Mark Shaw, Siobhan Tiernan, John F. O'Connell, David Warnock-Smith, Marina Efthymiou, (2021), Third party ancillary revenues in the airline sector: An exploratory study, *Journal of Air Transport Management*, Volume 90, January, 101936.

Chapter 4

Bernstein Vazirani and Deutsch Algorithm: Made Easy in Qiskit

R. Anitha[*], B. Vijayalaksmi[†], S. Anusooya[‡], V. Jean Shilpa[§] and C. Rishub

Department of Electronics and Communication Engineering, B. S. Abdur Rahman Crescent Institute of Science and Technology, Chennai, India

Abstract

Recent advances in Quantum Computing focus on the development of computer technology that depends on quantum physics. Extensive research work is under progress by researchers in various domains of Quantum computing, such as Quantum cryptography, Quantum annealing, Quantum Computer models, computational chemistry, etc., though we are in regular practice with classical computing, few systems of classical computing are still in the challenge which can be addressed by Quantum computing. Quantum computing uses quantum mechanics phenomena to make significant advances in computing to solve some problems. It relies on the foundation of the quantum mechanical theory of superposition and entanglement. This paper gives a better understanding of Quantum computing concepts and implementation of basic quantum algorithms like Bernstein Vazirani and Deutsch algorithm in the IBM, Quantum experience Qiskit tool which is an open-source tool that uses python to interface with quantum computers. Quantum computers developed by IBM can handle complicated issues that cannot be solved by efficient supercomputers.

[*] Corresponding Author's Email: r.anitha@crescent.education.
[†] Corresponding Author's Email: vijayalakshmi.b@crescent.education.
[‡] Corresponding Author's Email: anusooya@crescent.education.
[§] Corresponding Author's Email: jeanshilpa@crescent.education.

In: Frontiers in Quantum Computing: New Research
Editor: R. Anandan
ISBN: 978-1-68507-816-4
© 2022 Nova Science Publishers, Inc.

Keywords: quantum computing, quantum algorithm, qubits, entanglement, Qiskit, Bloch sphere

1. Introduction

Quantum Computing is not a new initiative. Once, in TEDx talk, Dr. Michelle Simmons, Founding Director of Silicon Quantum Computing Pvt. Ltd addressed that Quantum Computers will change the way of computing process. She demonstrated the power of quantum computers, which strengthens encryption and revolutionizes security. Finally, she ended by saying that there will be a huge race to build a quantum computer as it will be capable of tackling the problem in minutes which the classical computers fail to do.

Most of the researchers would think that the first quantum computer is yet to shine off, but IBM, Microsoft, and Google are ahead in a race representing the computing challenges. IBM has released Qiskit, an open-source tool for computing with the python framework. Similarly, Microsoft-Quantum Katas supports the development of Quantum Computing with Q# language. A simple example to define Quantum computing: Consider a case of dividing 3 pieces of cake for two kids. With a normal computer, for 3 bits, we were able to define only one of 8 solutions at one time — for example, 010. However, with a quantum computer, it's possible to define all 8 solutions at the same time.

1.1. Fundamentals of Quantum Computing

A classical computer works on bits whereas quantum computers work on qubits (quantum bits). Classical bits can exist in two states 0 and 1 but a qubit exists in the same states and also in linear combinations of both. These are termed superposition states.

A conceivable quantum state possessed by a qubit is represented by any 2D column vector of real or complex numbers with norm 1.

Thus, $\binom{\alpha}{\beta}$ defines a single qubit state, where "α"and "β" are the complex numbers satisfying the condition $|\alpha|^2+|\beta|^2 = 1$. The few examples of quantum state vectors defining qubits are $\binom{0}{1}, \binom{1}{0}, \binom{0.707}{0.707}, \binom{0.707}{-0.707}$. Among these states, the states $\binom{0}{1} and \binom{1}{0}$ are equivalent to the states of classical binary

digits, namely 0 and 1. The standard convention is to choose to be, $0 \equiv \binom{1}{0}$, $1 \equiv \binom{0}{1}$.

1.1.1. Transformations of Qubits Using the Bloch Sphere

The Bloch sphere model can also be used to visualize qubits in three dimensions. The Bloch sphere is a three-dimensional real-valued vector that can be used to describe a single-qubit quantum state shown in Figure 1.

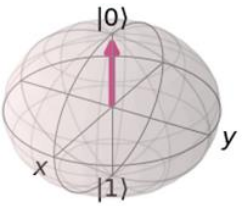


Figure 1. Bloch sphere.

The arrows in the above diagram indicate the direction of the quantum state vector, and each arrow transformation is considered as a rotation around one of the cardinal axes.

1.2. Simulating Polarisation of Light on a Quantum Computer

Quantum computing has been used to simulate a toy model of photon polarisation. Considering that the light exiting the first polarizer is horizontally polarised and if the second polarizer is placed with its pass-axis vertical, no light will be transmitted. So, these two states are said to be orthogonal and mapped two states to $|0\rangle$ and$1\rangle$ respectively (since these are also orthogonal). However, if the angle between the two orthogonal polarisation states is $\pi/2$, the angle on the Bloch sphere between $|0\rangle$ and $1\rangle$is π. Orientation of linearly polarised light is given as a superposition of the two basis vectors. Moreover, considering linearly polarised light, the relative phase is 0. Hence, considering the Bloch sphere passed through $|0\rangle$, $|+\rangle$, $|1\rangle$ and $|-\rangle$.

For this experiment, the parameterized circuit has been created, where the rotation angle is a parameter. Qiskit enables this using parameter(). Parameter β is considered aa s relative angle between the two polarizers. Given below Figure 2 is a quantum circuit where the qubit starts in the $|0\rangle$ state, the

equivalent of horizontally plane polarised light. The measurement basis is changed using a rotation gate with an angle of rotation in terms of the variable beta and measure on this basis. This basis change corresponds to rotating the second polariser by an angle β. Consider β to be positive in the counter-clockwise direction, as is the convention in Qiskit.

```
"# Qiskit Code
beta = Parameter('β')
qc = QuantumCircuit(1)
qc.ry(-2*beta, 0)
qc.measure_all()
qc.draw(output='mpl')"
```

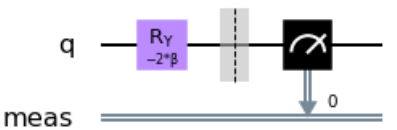


Figure 2. Quantum circuit equivalent of horizontally plane polarised light.

The part of the Bloch sphere visualized onto which mapping linearly polarised states has been done. As mentioned above, any state on the Bloch sphere can be created starting from |0⟩like so, $|\psi\rangle=(\phi)Ry(\theta)|0\rangle$. As described above, circle passing through |0⟩, |+⟩, |1⟩ and |−⟩ and the relative phase is 0, so ϕ=0 circle on the Bloch sphere. The relevant region is highlighted in Figure 3 below.

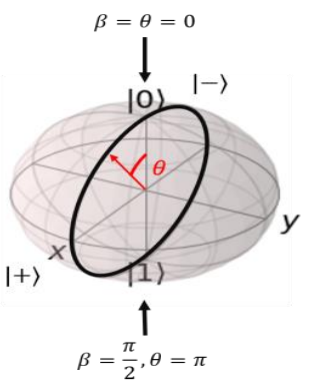


Figure 3. Bloch sphere mapping linearly polarised states.

Due to the mapping mentioned above, $\theta = 2\beta$, so a rotation of the polariser in the range $0 \leq \beta \leq \pi$ corresponds to $0 \leq \theta \leq 2\pi$. So the, polarisation state of a photon after the second polarizer (with the polariser at an angle β) can be constructed using

$|\beta\rangle = (2\beta)\,|0\rangle$; $(|\psi\rangle) = |\langle\beta|\psi\rangle|^2$

Taking the adjoint of the equation to prepare the state $|\beta\rangle$, $\langle\beta| = \langle 0|R^{\dagger}(2\beta) = \langle 0|Ry(-2\beta)$,

Where the last equality is evident from the form of the matrix given above.

With this, $(|\psi\rangle) = |\langle\beta|\psi\rangle|^2 = |\langle 0|R(-2\beta)|\psi\rangle|^2$

So, the appropriate gate to apply would be (-2β) and then perform a standard basis measurement. By rotating the second polariser through an angle of π i.e., $0 \leq \beta \leq \pi$and repeating the experiment for β values in this range, a photon can be measured that is passing through the second polariser.

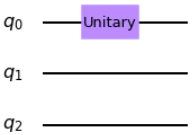
1.3. Entanglement and Teleportation Using Qiskit

According to the theory of quantum mechanics, teleportation means the move the quantum state of a photon to another without change. When two photons are said to be entangled, it shows the bond between the photons. If a photon is disturbed, the other will also immediately change, even if they are separated by a large distance. Its interesting fact is, observing the particles will change their quantum state that leading to a change of state of another particle that is entangled. Theoretically, even if they are light-years apart, observation of one state will change the other instantly without delay.

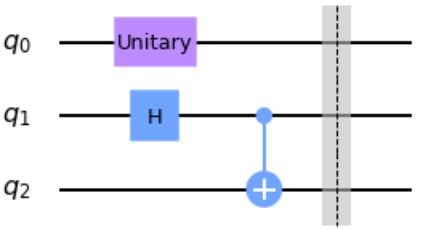
Teleportation will transfer the quantum information creating a quantum network. One benefit is that information can be transferred in a more secure form. Messages cannot be intercepted and read by a third party.

1.3.1. Quantum Teleportation Implementation in Qiskit

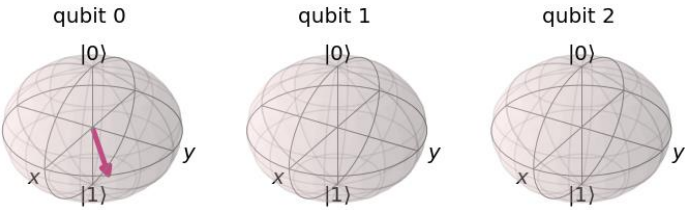
```
# Generate random unitary operator
u1q = random_unitary(2)
print(u1q.data)
state_prep = QuantumCircuit(3)
state_prep.append(u1q,[0])
state_prep.draw(output='mpl')
```



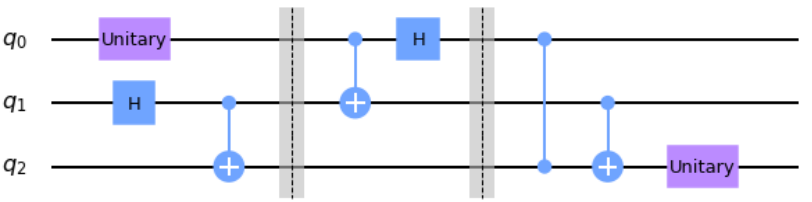

```
init_circ = state_prep.compose(bell_state_prep, qubits=[1,2])
init_circ.barrier()
init_circ.draw(output='mpl')
```



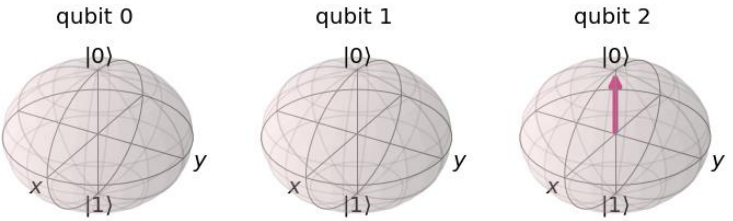

```
init_sv = Statevector.from_instruction(init_circ)
plot_bloch_multivector(init_sv)
```




```
tqc = init_circ.compose(bell_meas_circ, qubits=[0,1])
tqc.barrier()
tqc.cz(0,2)
tqc.cx(1,2)
tqc.append(u1q.adjoint(),[2])
tqc.draw(output='mpl')
```



```
tsv = Statevector.from_instruction(tqc)
plot_bloch_multivector(tsv)
```



1.4. Entanglement of Qubits

The entangled state of two qubits is defined by Bell states. Consider Person A and Person B spatially separated. The correlations between them can be easily explained with the help of quantum mechanics.

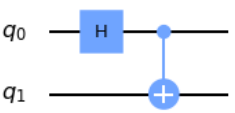
The Bell state $|\beta_{00}\rangle$or $|\Phi+\rangle$ is defined as $|\beta_{00}\rangle = \frac{|00\rangle+|11\rangle}{\sqrt{2}}$

Hadamard gate is used to transform $|00$ into its equivalent superposition state. The resultant output operates as a control input to the CNOT gate, inverting the second qubit when the first qubit is $|1\rangle$. Thus, CNOT gate transforms the second qubit as $\frac{|00\rangle+|11\rangle}{\sqrt{2}}$

The quantum circuit is created with the following code in the Qiskit tool,

```
b_s = QuantumCircuit(2)
b_s.h(0)
b_s.cx(0,1)
b_s.draw(output='mpl')
```

The above code defines the preparation of quantum circuits in the Qiskit tool.

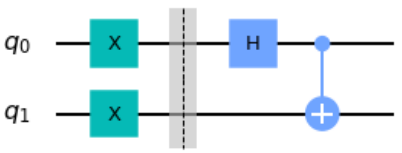


1.4.1. Bell States and Composing Quantum Circuits

Often in quantum algorithms, the circuit for an algorithm is composed of smaller repeated blocks with a particular function. To illustrate this, all the four Bell states have been defined.

Creating a Bell state uses the same gates as the |β00⟩ state, but with some gates pre-pended before this block.

```
def bell_circuit(state):
bell = QC(2)
for i in range(2):
if (state[i] == '1'):
bell.x(i)
bell.barrier()
bell.compose(bell_state_prep, [0,1], inplace=True)
return bell
bij_circ = create_bell('11')
bij_circ.draw(output='mpl')
```



1.4.2. Measurement in the Bell Basis

Bell basis can be measured using standard basis measurements and the inverse unitary operation. In the same method of bell state preparation, the bell basis can be measured and its circuit is shown in Figure 4.

```
# Create one of the four Bell states
bij_circuit = create_bell_circuit('01')
bij_circuit.barrier()

# Invert the Bell state preparation circuit as a measuring device
bell = bell_state_prep.inverse()

# Join everything together
total_bell_circuit = bij_circuit.compose(bell_meas_circ, [0,1])
```

```
# Add measurement and draw the circuit
total_bell_circuit.measure_all()
total_bell_circuit.draw(output='mpl')
```

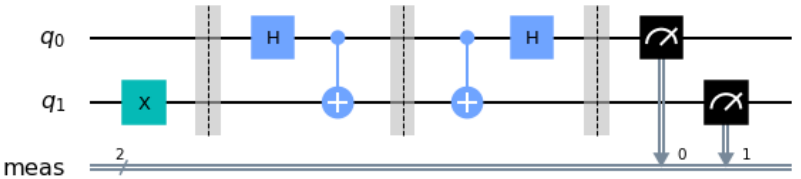


Figure 4. Bell circuit.

2. Deutsch Algorithm

Deutsch algorithm is one of the oldest algorithms which show Quantum Supremacy over classical algorithms [1-2]. The problem is that a function can take either 0 or 1 but cannot be defined as the actual state.

Assuming two functions and hence 4 different states exist as shown in Table 1.

Table 1. Outcomes for the two functions

φf(0)	φf(1)	function
|0⟩	|0⟩	"Constant"
|0⟩	|1⟩	"Balanced"
|1⟩	|0⟩	"Balanced"
|1⟩	|1⟩	"Constant"

Since, only two types of input existφ0 and φ1, the output is going to be from φf(0) and φf(1). From the tabulation, the possible outputs from the given two functions are, if both the output from both the functions zeros or ones they are a constant function or if either of the outputs is 0 or 1 are considered as a balanced function [3-4].

Four possible outcomes for the function is given below φf:

- Both the inputs φ0, φ1 mapped to the output φ0 = φf(0) = φf(1).
- Both the inputs φ0, φ1 mapped to the output φ1 = φf(0) = φf(1).
- Input pass through φf unaffected φf(0) = φ0 and φf(1) = φ1
- Inputs are exchanged after passing through the function input φf(1) = φ0 and φ (0) = φ1.

Consider, the first two cases of the function to be a constant function as it remains the same way, even after a function has been insertedand the last two cases of the quantum function behave as a balanced function.

2.1. What Does a Classical Algorithm Do?

Simple, if then else statement can be used to determine whether is it going to be a constant or a balanced function.

The code given below can help simplify how a classical algorithm works,

```
if g(0) = 0;
        if g(1) = 0:
                print ("Constant Function")
        else:
                print ("Balanced Function")
else:
        If g(1) = 0:
                print ("Balanced Function")
        else:
                print ("Constant Function")
```

where the “g” function is the input qubit function.

Quantum algorithms exploit the superposition of different states in Quantum Machines to define this problem efficiently in just one step [5].

2.2. Quantum Representation of the above Problem

For simplicity, assume that the given Quantum function maps input for the given state:

φ f(1) = φ0 and φ f(0) = φ1

Assume the bits as vectors,

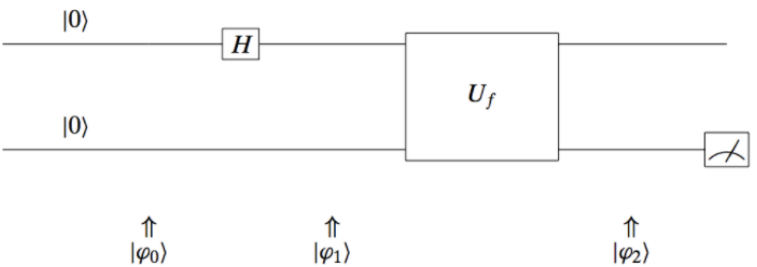
$|0\rangle=\begin{pmatrix}0\\1\end{pmatrix}$ and $|1\rangle=\begin{pmatrix}1\\0\end{pmatrix}$ can be represented as, $U_f=\begin{bmatrix}0 & 1\\1 & 0\end{bmatrix}$

Initially, the system is in the given state:

$$\varphi|0\rangle = 00\rangle = \begin{bmatrix}1\\0\\0\\0\end{bmatrix}$$

By concatenating two 1-qubit states, a qubit can be represented as follows,

$$\begin{bmatrix}00\\01\\10\\11\end{bmatrix}$$



The top input is transformed by the Hadamard operator,

$$|\varphi_1\rangle = |H\cdot 0,\ 0\rangle = \left|\frac{|0\rangle+|1\rangle}{\sqrt{2}},\ 0\right\rangle = \frac{|00\rangle+|10\rangle}{\sqrt{2}}$$

Then, by applying the U_f transformation: U_f can be applied to the top of the input and to define the output at the bottom, output $|y \oplus f(x)\rangle$ is used, therefore φ2 becomes:

$$|\varphi_2\rangle\left(=\frac{|x,y\oplus f(x)\rangle+|x,y\oplus f(x)\rangle}{\sqrt{2}}\right)=\frac{|0,0\oplus f(0)\rangle+|1,0\oplus f(1)\rangle}{\sqrt{2}}$$
$$=\frac{|0,0\oplus 1\rangle+|1,0\oplus 0\rangle}{\sqrt{2}}=\frac{|01\rangle+|10\rangle}{\sqrt{2}}$$

From the measurement, it's possible to have a 50% probability of finding it states similarly for most of the bottom qubits.

2.3. Quantum Representation: Deutsch Algorithm

The Deutsch circuit is shown in Figure 5.

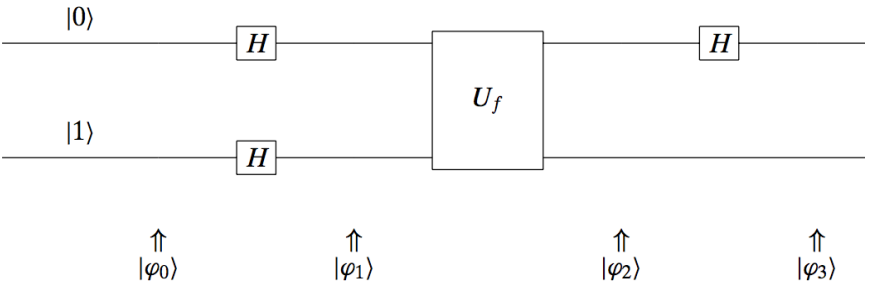


Figure 5. Deutsch circuit.

By observing the above circuit, the known states are defined $\varphi|0\rangle = |0,1\rangle$ [6]. When both the inputs pass through the Hadamard Gate the resulting equation would be as follows,

$$|\varphi_1\rangle = |H \cdot 0, H \cdot 1\rangle = \frac{|0\rangle + |1\rangle}{\sqrt{2}} \otimes \frac{|0\rangle - |1\rangle}{\sqrt{2}}.$$

By the definition of φf, the output is rewritten as:

$$(-1)^{f(x)} \frac{|x,0\rangle - |x,1\rangle}{\sqrt{2}} = (-1)^{f(x)}|x\rangle \otimes \frac{|0\rangle - |1\rangle}{\sqrt{2}}.$$

The change in function behaviour is observed by taking different cases of φf:

φf(0) = 1φ and φf(1) = 0φ,

$$|\varphi_2\rangle = \frac{(-1)^1|0\rangle + (-1)^0|1\rangle}{\sqrt{2}} \otimes \frac{|0\rangle - |1\rangle}{\sqrt{2}} = (-1)\frac{|0\rangle - |1\rangle}{\sqrt{2}} \otimes \frac{|0\rangle - |1\rangle}{\sqrt{2}}$$

For all the four input cases framed, the summarized output equation is given below:

$$|\varphi_2\rangle = \begin{cases} (\pm 1)\frac{|0\rangle+|1\rangle}{\sqrt{2}} \otimes \frac{|0\rangle-|1\rangle}{\sqrt{2}}, & \text{if } f \text{ is constant,} \\ (\pm 1)\frac{|0\rangle-|1\rangle}{\sqrt{2}} \otimes \frac{|0\rangle-|1\rangle}{\sqrt{2}}, & \text{if } f \text{ is balanced.} \end{cases}$$

The resulting output equation is given as follows,

$$|\varphi_3\rangle = \begin{cases} (\pm 1)|0\rangle \otimes \frac{|0\rangle-|1\rangle}{\sqrt{2}}, & \text{if } f \text{ is constant,} \\ (\pm 1)|1\rangle \otimes \frac{|0\rangle-|1\rangle}{\sqrt{2}}, & \text{if } f \text{ is balanced.} \end{cases}$$

2.4. Qiskit Implementation

A Deutsch algorithm with a three-bit function that has both constant and balanced oracles is discussed below.

2.4.1. Constant Oracle

Initialize a constant oracle and at random, place the output qubit as"0" or "1":

"# Qiskit code:

```
c_o = QuantumCircuit(n+1)
out_put = np.random.randint(2)
if out_put == 1:
c_o.x(n)
c_o.draw()"
```

2.4.2. Balanced Oracle

A balanced oracle is created by applying a CNOT operation on each input qubit which is a control and the output bit is considered as the target bit. Wrap the controls in X-gates such that the input states can be changed. In-circuit, an X-gate is used if and only if the digit in b_str is "1."

“# Qiskit code:

b_o = QuantumCircuit(n+1)
*b_str =“101” #*binary string of length ‘n’ that specifies which controls to wrap

 # Place X-gates
for qubit in range(len(b_str)):
if b_str[qubit] ==‘1’:
b_o.x(qubit)

 # Use barrier as divider
b_o.barrier()

 # Controlled-NOT gates
for qubit in range(n):
b_o.cx(qubit, n)

b_o.barrier()

 # Place X-gates
for qubit in range(len(b_str)):
if b_str[qubit] ==‘1’:
b_o.x(qubit)

 # Show the oracle
b_o.draw()”

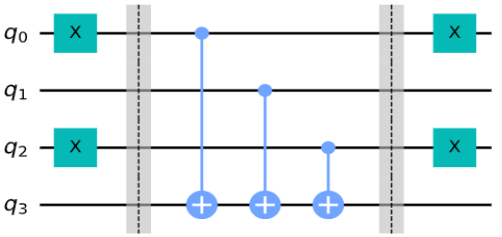


Figure 6. Oracle circuit.

Deutsch Algorithm:

- The input qubitsare initializedin the state |+⟩ and the output qubitis initialized in the state |−⟩.

- Apply the oracle
- H-gates are applied on the n-input qubits and measure the states [7].

“Qiskit code:

```
deustch_circuit = QuantumCircuit(n+1, n)

# Apply H-gates
for qubit in range(n):
deustch_circuit .h(qubit)

# set qubit in thestate |−⟩.
deustch_circuit .x(n)
deustch_circuit .h(n)

# Add oracle
deustch_circuit += b_o

# Repeat H-gates
for qubit in range(n):
deustch_circuit .h(qubit)
deustch_circuit .barrier()

# Measurement of states
for i in range(n):
deustch_circuit .measure(i, i)

# To display deustch_circuit
deustch_circuit .draw()”
```

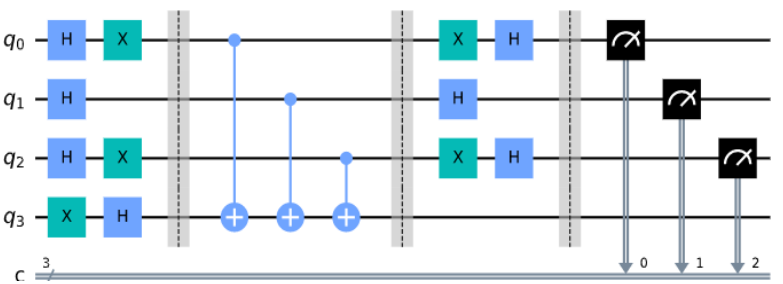


Figure 7. Deutsch circuit.

3. Bernstein-Vazirani Algorithm

It is possible to differentiate between constant & balanced Boolean functions using the Deutch-Jozsa technique. The Deutsch-Jozsa algorithm has an extension that allows extracting information about a different class of functions.

A n-bit Boolean function $f(x)$ is called linear if it can be written as the bitwise product of a particular n-bit binary string a and the function variable x (which is also a binary string of length n), i.e., linear functions can be written as "$f(x)=a.x \pmod 2$"

For general n-qubit computational basis state, the Hadamard transform has the following effect,

$$H^{\otimes n}|a\rangle = \frac{1}{2^{n/2}} \sum_{x=0}^{n-1} (-1)^{a \cdot x} |x\rangle$$

Apply $H^{\otimes n}$ to both sides of the above equation and get (after flipping sides),

$$H^{\otimes n} \left(\frac{1}{2^{n/2}} \sum_{x=0}^{n-1} (-1)^{a \cdot x} |x\rangle \right) = |a\rangle$$

The term inside the brackets represents the function if an equal superposition state is passed through a phase oracle for the function, $f(x) = a.x \pmod 2$ as mentioned below:

$$\frac{1}{2^{n/2}} \sum_{x=0}^{n-1} |x\rangle \xrightarrow{U_f} \frac{1}{2^{n/2}} \sum_{x=0}^{n-1} (-1)^{a \cdot x} |x\rangle$$

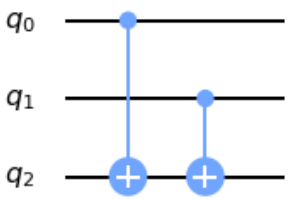
The Bernstein-Vazirani algorithm uses all the things discussed above. Given an oracle for a function that we know is linear, we can find the binary string corresponding to the linear function. The steps of the algorithm are shown in the equation below and then described in words.

$$|0^{\otimes n}\rangle \xrightarrow{H^{\otimes n}} \frac{1}{2^{n/2}} \sum_{x=0}^{n-1} |x\rangle \xrightarrow{U_f} \frac{1}{2^{n/2}} \sum_{x=0}^{n-1} (-1)^{a \cdot x} |x\rangle \xrightarrow{H^{\otimes n}} |a\rangle$$

Considering, the Boolean function $f(x) = (x1' \wedge x0) \vee (x1 \wedge x0')$ as a linear function. The above-mentioned expression is implemented as the quantum bit oracle for this Boolean function.

A QuantumCircuit() with 3 qubits is provided below.
q0 and q1 are the input qubits for the variables x_0 and x_1 respectively.
q 2 is the output qubit and stores the value of the final Boolean function expression

```
b_v_circ = QuantumCircuit(3)
b_v_circ.cx(0,2)
b_v_circ.cx(1,2)
b_v_circ.draw('mpl')
```



Using the bit oracle, the Quantum circuit for the Bernstein-Vazirani algorithm has been discussed below:

The steps for the algorithm are as follows:

1. Start with"(*n*+1)"qubits in the |0⟩ state. Here "*n"* =2. The first two qubits q0 and q1 will serve as input to the oracle. The extra qubit is used for the oracle output. Since we need a phase oracle, add gates to prepare the state |-⟩ in this qubit (q2).
2. Apply an H gate to all the input qubits.
3. Apply the oracle U_f
4. Apply a gate to all the input qubits.
5. Measure the *n* input qubits. If the function corresponding to Uis linear, the final state measured will be the binary string *a.*

```
b_v = QuantumCircuit(3,2)
b_v.h(2)
b_v.z(2)
b_v.h(range(2))
b_v.barrier()
```

```
b_v.extend(bv_oracle)
b_v.barrier()
b_v.h(range(3))
b_v.measure([0,1],[0,1]) # ([list of qubits], [list of bits])
b_v.draw(output='mpl')
```

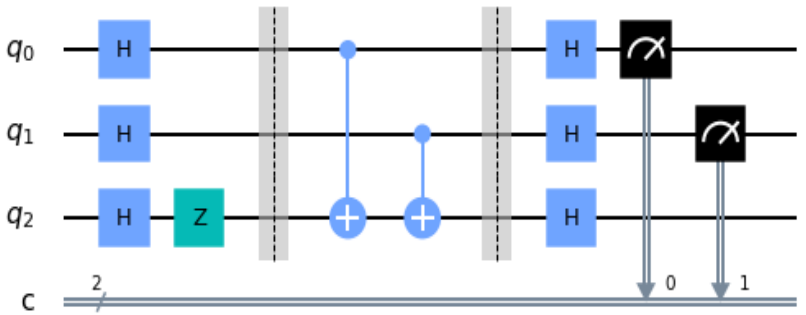


Figure 8. Bernstein-Vazirani circuit.

Conclusion

As the possibility of commercial quantum computers becomes increasingly feasible in recent years, quantum algorithm research has become the focus of attention. Ordinarily, intractable computation problems could now be addressed very efficiently with non-classical approaches, thanks to quantum computing's great parallelism in Hilbert space. To take advantage of parallelism, innovative quantum algorithms are needed, allowing efficient quantum oracles to be tailored to individual processing requirements. As a result, quantum algorithms and their applications are just as crucial as the quantum computer technology in the race for quantum supremacy. The basic concepts of quantum processing are covered in this article, as well as some key quantum algorithms.

References

[1] Samantha Buck, Robin Coleman, and Hayk Sargsyan, "*Continuous Variable Quantum Algorithms: an Introduction*," arXiv preprint arXiv:2107.02151, 2021.

[2] Carmen G. Almudever, Lingling Lao, Robert Wille, Gian Giacomo Guerreschi, "Realizing Quantum Algorithms on Real Quantum Computing Devices,Published in 2020 Design, *Automation & Test in Europe Conference & Exhibition* (DATE).

[3] David Deutsch, Richard Jozsa, "Rapid solutions of problems by quantum computation," *Proceedings of the Royal Society of London A.* 439 (1907): 553–558, doi:10.1098/rspa.1992.0167. S2CID121702767, 1992.

[4] R. Cleve, A. Ekert, C. Macchiavello, and M. Mosca, "Quantum algorithms revisited," *Proceedings of the Royal Society of London A.* 454 (1969): 339–354, arXiv:quant-ph/9708016, 1998.

[5] Simon and Daniel, "On *the Power of Quantum Computation," Proceedings of the 35th Annual Symposium on Foundations of Computer Science*," 116–123. doi:10.1109/SFCS.1994.365701. ISBN0-8186-6580-7, 1994.

[6] Johansson, Larsson et.al, "Efficient classical simulation of the Deutsch–Jozsa and Simon's algorithms," *Quantum Information Processing*, vol. 16(9), pp.233. arXiv: 1508.05027, 2017.

[7] David Deutsch, "Quantum Theory, the Church-Turing Principle and the Universal Quantum Computer," *Proceedings of the Royal Society of London* A.400(1818): 97–117,doi:10.1098/rspa.1985.0070, 1985.

[8] *Qiskit Deutsch-Jozsa Algorithm and Implementation in Qiskit* https://qiskit.org/ textbook/ch-algorithms/deutsch-jozsa.html.

Chapter 5

Quantum Aided Deep Learning Framework for Motif Structure Prediction

S. Nithinsha[1,*] and D. Vinod[2,†]

[1]Department of CSE, Saveetha School of Engineering, SIMATS, Chennai, India
[2]Department of Cloud and Edge Analytics,
Institute of Computer Science and Engineering,
Saveetha School of Engineering, Chennai, India

Abstract

New drug discovery and development requires an in-depth study of technological, scientific, and managerial factors of protein sequences. Failing this leads to rejection in drug approvals. Conventional reviews about motifs are very manual, which leads to partial knowledge of molecules and their interactions. On the other hand, computational intelligence techniques using deep learning algorithms result in a net gain of information from each drug discovery, which discovers efficient drugs faster and easier. This research focuses mainly on leveraging sequence, structure, and functions of proteins to identify motifs, usage motif embeddings for perfect vector formation, and deep learning models to determine the motif patterns. This research work proposes a deep learning framework for DNA binding motif identification using motif embeddings and applying a deep learning network named Deep Convolutional Networks. Data analytics play a very important role in analyzing unstructured data. Structure classification of protein sequence is a hectic task and now deep learning techniques are used to solve this problem. The proposed approach uses NLP based data representation together with a Deep Learning approach to efficiently predict motifs.

[*] Corresponding Author's Email: ayshaas2017@gmail.com.
[†] Corresponding Author's Email: dvinopaul@gmail.com.

In: Frontiers in Quantum Computing: New Research
Editor: R. Anandan
ISBN: 978-1-68507-816-4

© 2022 Nova Science Publishers, Inc.

Keywords: deep neural networks, motif, prediction, convolution networks, bioinformatics

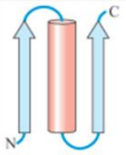
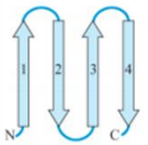
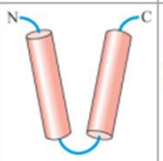
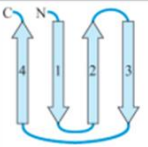
1. Introduction

Conventional studies about motifs are very manual which leads to partial knowledge of molecules and its interactions. On the other hand, computational intelligence techniques using deep learning algorithms results in a net gain of information from each drug discovery, which in turn makes the discovery of efficient drugs faster and easier. In the past decade, technological advances paved the way to explore and research more on bioinformatics. Among various researches, genomics and proteomics lead the topmost research topics [1]. A study of proteomics data showed rapid growth, and more than 2,30,000 protein structures have been identified, which are available in various protein data banks in multiple formats [2]. The number is still growing exponentially, leading to different research areas, including protein structure prediction, motif identification, motif clustering, and motif formation prediction [3]. Today, a vast number of pharmacists and scientists working in the field of proteomics are leveraging their knowledge to formulate principles and rules on how to construct motifs, motif designs, motif identification, and factors that affect the formation of motifs [4]. This research aims to understand small organic molecules to trigger or inhibit a specific function of motif or disrupt protein interactions, which helps treat a disease.

2. Background

2.1. Motifs

Proteins perform a critical function to maintain our health and wellbeing. Motifs or super secondary structures are specific combinations and more than one secondary system, mostly with defined connectivity. Protein sequences and structures contain patterns of similar amino acid composition with distinct functions [5]. These patterns are nothing but motifs with a mixture of secondary and tertiary structure elements to form the final protein with complex functionality. Figure 1 shows a few commonly available motifs.

Motif	Schematic	Description	Comments	Examples
βαβ		An α helix connects parallel β strands in β sheets. Hydrophobic surfaces on the helix and β sheet interact.	Residues in the first loop (C-terminal end of the β strand) often contribute to the active site in enzymes.	triose phosphate isomerase (a glycolytic enzyme) (pdb file 1tim)
β meander		Antiparallel β sheets are linked sequentially by short loops or hairpins consisting of two, three or four residues.	Basic antiparallel β pleated sheet.	
helix–turn–helix		In the simplest arrangement, two helices lie antiparallel, connected by a short loop.	Energetically favourable interactions between side-chains are accommodated by the relative positioning of the helices (usually at ~20° to each other).	Rop (an RNA binding protein) (pdb file 1rop)
Greek key		Antiparallel β sheet with longer loop connections between some strands.		gamma crystallin (a protein of the eye lens) (pdb file 1gcs)

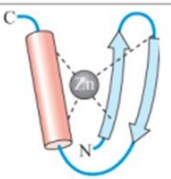
Motif	Schematic	Description	Comments	Examples
coiled-coil		Two α helices are wrapped around each other. Strips of hydrophobic side-chains along the length of each helix interact with each other.		myosin (a motor protein); α-keratin (a structural protein, e.g. in skin and hair)
zinc finger		An α helix and two short antiparallel β strands are held together by a zinc ion, which forms coordinate bonds with side-chains in the polypeptide.		Xfin (*Xenopus* DNA binding protein with a role in embryogenesis) (pdb file 1znf; single zinc finger from this protein)

Figure 1. Commonly occurring motif types.

It is more important to identify the motifs, mainly DNA binding proteins, for safe and successful therapeutics for efficient drug discovery. An efficient approach is essential to identify motifs for accurate drug discoveries. Extraction of features from biomedical data is one of the crucial steps in any learning models. This step focus on ways to extracts interpretable information from biomedical data and is proportional to the performance of machine learning algorithms [6]. Various approaches are being followed across researches, and these types can be broadly grouped under 2 categories 1) Supervised 2) Unsupervised.

1. Supervised - Sequence Encoding Schemes

For structure prediction, some common sequence encoding schemes which extracts features such as PSSM profiles, amino acid factors, physio-chemical properties, hydrophobicity, surface tension, polarizability, polarity, side chain mass, charge, protein sequences and solvent

2. Unsupervised - Distributed Representation

An unsupervised data representation for biological sequences have been proposed [7] named as BioVec. BioVec (bio vectors) are ProtVec which are vectors extracted using word embeddings which can be applied for various problem statements like domain extraction, protein visualization, structure prediction, protein family classification. This approach characterizes physical and chemical properties of bio-sequences and embedded in an n-dimensional vector. One of the main advantages of word embeddings is that, the vectors generated could be reused for any kind of biological problems. Deep learning in bioinformatics has lot of benefits from these prebuilt vectors. This is also being considered as pre-trained models for various deep learning algorithms.

Unstructured text data is analogy with protein sequence data and so NLP techniques are widely used in proteomics. Each letter in motif signifies an amino acid with which these NLP techniques converts amino acid into text form without losing the position information of motifs. These techniques do not require traditional feature engineering with an advantage of giving good accuracy for any deep learning models. Inappropriate features lead to poor performance of learning models which degrades the accuracy of deep learning network. The basic idea behind this vectorization technique is that the meaning of any term in a sentence is characterized by its neighboring word. i. e. by its

context. Skip Gram and CBOW are two variants of this shallow network which uses different type of input for generating the vectors.

- Skip-Gram: The skip gram model consists of an input layer, embedding (hidden) layer, and output (prediction) layer. This is an unsupervised technique to predict the context word for a given target word. The results are the word embedding based vectors generated. By using skip gram model, the probability is maximized by adjusting network parameters.
- Continuous Bag of Words: CBOW model consists of an input layer, embedding (hidden) layer, and output (prediction) layer as same as Skip gram. This is also an unsupervised technique to predict the target word based on context words.

In this study we have represented our motif's using 2 way of vector representation. First way is to create custom embeddings using input motifs. Protein structure prediction algorithms are classified into numerous levels. They are ranging from one dimensional to four dimensional protein structure prediction.

2.1.1. Evolution of Motif Classifiers

The prediction of motif's has a long history and various classifiers have been studied [16]. Fifth generation methods mainly rely on,

1. Template Free - Multiple sequence alignments of target sequences.
2. Template based.

Deep Learning algorithms are showing promising results in data analytics arena along with distributed processing [8]. We have aimed for accuracy and performance improvement by implementing these techniques in our work. Among these, Deep Learning classifiers consists of cascaded sequence of multiple layers with the capability to perform non-linear processing, automatic feature extraction and transformation thereby deriving additional high level of abstracted features from low level input features performs well [9]. Also, Deep Learning exhibits few properties of representation learning wherein these models accepts raw data in vectored form like word embedding's, N-hot vectors or image pixel values and the models learn features from raw data automatically as depicted in [10].

Feature extraction using handheld techniques degrade the performance of Deep learning models. Word embeddings-based motif vectors provides a great help which works best when implemented in deep learning network. Deep learning is one of the representations learning methods which can discover and learn hierarchical representations of data. Several deep learning architectures exists such as Deep Neural Network, Stacked Auto Encoders, Convolutional Neural Network, Recurrent Neural Network and Deep Belief Network. This study uses convolutional neural network which is trained on the basis of motif embeddings [11].

Motif embeddings after learning can be visualized as points in a multi-dimensional embedding space or one-dimension sequence logos. We displayed that, from multi-dimensional space evolutionary distant homologous sequences can be highlighted, and comparison of homologous motifs can be visualized using tSNE [12]. We define two metrics to compare the search results with current linear sequence motif based methods. Convolutional neural network departs from the motif extraction status quo, namely multiple sequence alignment algorithms, which require fine hand tuning of motif search parameters, and performs a search automatically tuned to produce results satisfying a given optimization criterion, such as maximum functional or structural similarity. We investigate the generative ability of convolutional neural network and develop computational methods that allow the directed evolution of proteins towards altered or novel functions.

2.1.2. Importance of Structure Prediction

In the field of bioinformatics, researchers are facing a challenging task in designing and developing new strategies for data processing and drug development. The analysis of protein sequences and prediction of their structures poses daunting and intellectually exciting problems for computational science [13]. The growth of sequence databases and the need for fast detection of similarities between sequences are driving researchers into data compression, string-matching algorithms, network and database theory, and design. A distributed approach would be a viable option for massive quantity of protein data and also for forecasting the motif structures of huge size.

Structure prediction is necessary because the knowledge of protein structure (and hence function)is crucial to an understanding of biology and to progress in biotechnology and medicine, and because it is currently the only (relatively) quick and inexpensive technique to discover the native structure of a newly discovered protein. The deduction of protein structures from X-ray

crystallographic and NMR imaging data is difficult, slow, and costly [14]. A potentially useful application of scanning tunnelling microscopy to the problem is probably not feasible for at least several years. About 1200 to 2000 unique protein and poly-peptide structures have been solved, while over 200,000 sequences have been archived into the global databases. The complete understanding of the protein structures is very much essential before undergoing the Motifs prediction model process.

Protein Structures

Proteins are the most generous natural particles in human being, assuming significant parts in all parts of cell construction and capacity. Proteins are biopolymers of acids, so named in light of the fact that the amino gathering is clung to the carbon iota, close to the carbonyl gathering. The physical and compound properties of a protein are controlled by its principal amino acids. The specific amino corrosive sub-units are combined by amide bonds called peptide bonds. Proteins are comprised of building blocks assembled amino acids which shapes a huge particle that assumes basic parts in the body. It is needed for construction, capacity, and map line of tissues inside the body [15]. There are a huge number of fluctuated sorts of proteins, like chemicals, hormonal, antibodies, and primary.

Proteins are the premise, all things considered. They partake taking all things together natural cycles within the human body. All proteins are polymers of amino acids for example amino acids are the essential structure squares of protein. Each protein is a remarkable string of the twenty amino acids. They vary between each other in the number and grouping of amino acids. Each protein has a specific shape and compound characteristics based on the amount and sequence of amino acids. The succession of amino acids of a protein addresses a natural data unit. These data make exceptional construction of proteins that decides their movements (substance properties) in the creature.

Proteins are essential for living organisms because they include diverse amino acids. The names of these amino acids are written in a 3-letter code with significant variations. These amino acids are assembled in accordance with several conditions comprising hydrophobicity, size, aromaticity, or charge. Fundamentally, proteins have four distinct constructions as shown in Figure 2 namely, primary, secondary, quaternary and tertiary structures [16]. The primary structure is the foundation for all kinds. Secondary structures combine to form Motifs like β-hairpin, α-hairpin, helix turn helix, β- α- β. A tertiary

structure is made up of secondary structures that combine to form polypeptide chains. Tertiary structure gives rise to quaternary structure

- Primary Structures: The essential construction alludes to the exceptional amino corrosive arrangement of the polypeptide series. The essential design is molded by covalent or peptide bonds.
- Secondary Structures: Secondary design alludes to profoundly ordinary neighborhood substructures. These optional designs are characterized by examples of hydrogen connections amongst the primary string peptide gatherings. The three optional constructions are: alpha helix, beta sheets and loop that are affected by the characteristics of every amino corrosive. The super secondary structures (Motifs) are shown in Figure 2.
- Tertiary Structures: Tertiary structure alludes to 3D design of a solitary protein atom. The alpha-helices and beta-sheets are collapsed into a conservative globule. The3D constructions are liable for utilitarian attributes of proteins.
- Quaternary Structures: Quaternary structures are made out of at least two sub-units of tertiary constructions.

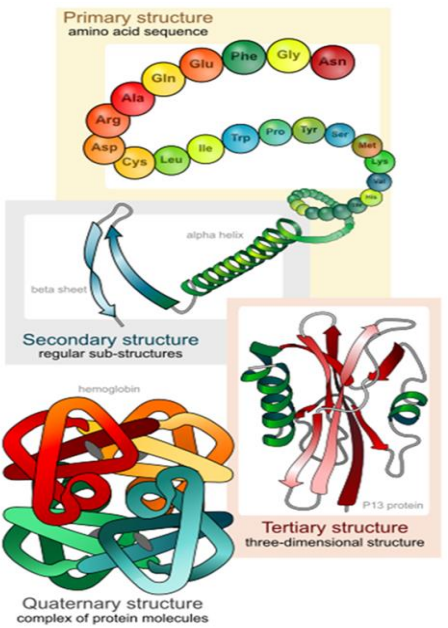


Figure 2. Structures of protein.

2.2. Need for Current Research Work

Most researchers are focused on identifying regulatory elements with genome sequence of motif structure prediction. Hence for the prediction of super secondary structure, a focus on structural and sequence element is required. Conventional studies about motifs are very manual which leads to partial knowledge of molecules and its interactions. Computational Intelligence (CI) techniques using Deep Learning (DL) algorithms results in a net gain of information from each drug discovery, which in turn makes the discovery of efficient drugs faster and easier.

Many studies have investigated the prediction of structured data in motif super secondary structure. They have not explored all types of motifs. Handmade feature extraction is used by earlier approaches for the knowledge extraction from amino acid sequence. Here automatic feature extraction is considered for improving the accuracy. The recurrent secondary structures are coiled coil, Helix Turn Helix and Zinc finger motif, β sheet and coils and also motifs.

One of the prominent research studies titled *Ab initio* super secondary structure prediction method is used to produce tertiary structure forecasts that demands in accordance with discovery of proteins rapidly. Though recent trends and developments have attained more accurate results than previous methods of SSS prediction, there was the delimiter 80% accuracy. But it is crossed that barrier level. Neural networks have been considered as an important part. This approach is a semi-supervised algorithm. There is no study for prediction of unstructured data. In this aspect, the vector using Natural Language Processing (NLP) approach is used in a distributed SPARK environment for improving accuracy and performance.

2.3. Map Reduce Framework

The Figure 3 illustrates the block diagram of Apache Map Reduce Framework. Apache Hadoop is an open source development for dealing with and scattering trustworthy records. The Apache Hadoop programming library is a framework that considered into account the circulated preparing of enormous informational collections from corner to corner clusters of Personal Computers consumes direct forward programming models. It is anticipated to increase from single workers to a huge number of machines, every contribution neighborhood calculation and capacity. Maybe that depends on

equipment to express high availability, the actual library is anticipated to recognize and manage dissatisfactions at the application layer, so passing on a particularly open assistance on top of a lot of Personal Computers, all of which may be leaning towards dissatisfactions.

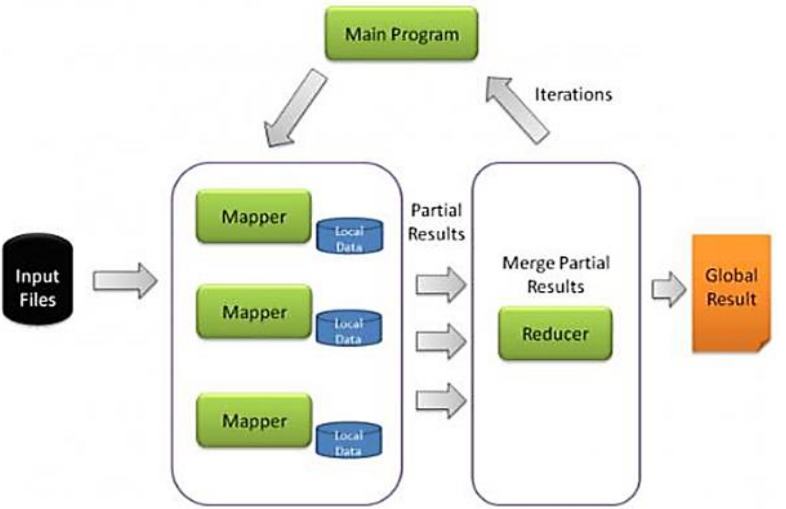


Figure 3. Apache map reduce framework – block diagram.

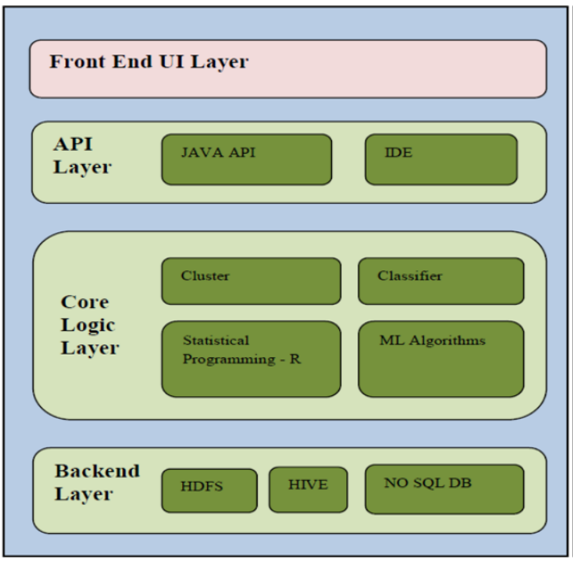


Figure 4. Proposed architecture.

The item obtained from it facilitates in the appropriated treatment of enormous educational assortments. They are expected to increase from a single worker to a large number of machine present neighborhood estimation and limit. Artificial neural network is a directed AI strategy for protein structure prediction. ANN generates output for explicit information design. Numerous neural organization structures, such as multi-facet perceptron, spiral premise work, and back engendering network, are available in writing. Computers have always been a particularly effective technique of dealing with massive volumes of data. A platform like the Big Data Framework delivers effective processing capacity, bandwidth, storage, security, and system dependability. PSSP is the method of recognizing and forecasting the structures in a protein sequence, and this thesis focused on leveraging Hadoop and Map Reduce techniques to apply the ML method. The proposed architecture is depicted in Figure 4.

Map Reduce measures enormous volume of huge information by utilizing equal and disseminated figuring on huge groups of frameworks, hubs, or in mists, which comprise of an expert hub and numerous laborer hubs. MapReduce depends on the utilitarian programming idea that comprises of the map, lessen and join capacities. As Hadoop MapReduce executes on the plate, that is generally sluggish & costly because of the great Input/output activities. Interestingly, as Spark MapReduce execution is acted in memory of figuring hubs on the group, it gives a quicker & further financially savvy method of carrying out versatile equal and appropriated calculations for large information examination.

2.3.1. API Layer

In the above proposed architecture, under the API Layer, The query protein pairings are used to dynamically parse PDB files using the JAVA Map Reduce application. Every coordinating file is broken down into parts and saved to HDFS. Integrated Development Environment (IDE) is actually a software application that offers complete facilities to software developers. This environment brings them necessity tools to write or test the software application and also modify or write a new code while compiling, debugging and executing it at the same place. IDE can be a part of one or more compatible applications or standalone application. The flow of data for program development is depicted in Figure 5:

R is a statistical analysis, visual representation, and reporting computer language and software environment. The R Development Core Team is actively working on R, which was founded by Ross Ihaka and Robert

Gentleman at the University of Auckland in New Zealand. R is available as Public License, and pre-compiled binary versions are available for Linux, Mac OS, Windows, and Mac. The R programming language was called by the 1st letter of the two authors' labels (Robert Gentleman and Ross Ihaka). R is an interpreted computer language that promises to provide branching, looping, and modular programming using functions. R Programming is seen in Figure 6. It enables for seamless interaction with C, C++, Net, Python, and FORTRAN functions. Message Code and bug reports have been sent in by a large number of people who have contributed to R.

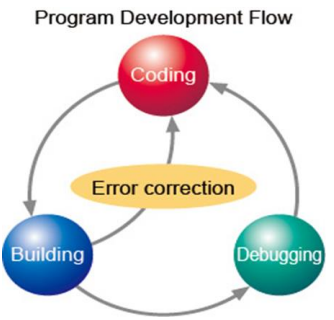


Figure 5. Program development flow.

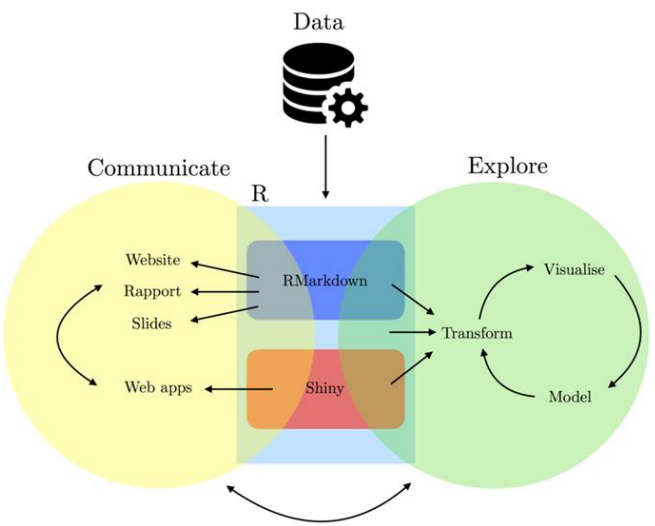


Figure 6. Working of R programming.

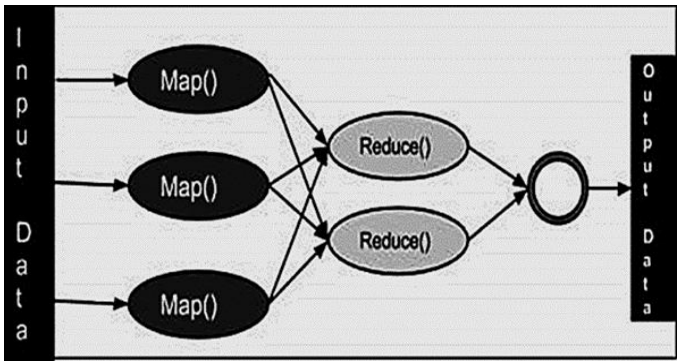


Figure 7. Mapper and reducer.

During the mapping phase, a mapper with such a primary key and a value runs its algorithm in the core logic layer. It processes the partition data and generates the intermediate result. It works as a cluster. Each of the proteins chosen can still have up to five mappers. Before handing the data to the Reducer, each Mapper might utilize the Mahout package to categorizes it. During the reduction step, a reducer programme is executed to gather intermediate results and report them to HDFS. It must function as a classifier. The reducer will get input from each of the five Mappers after an appropriate Shuffle and Sort phase. This Map Reduce phase helps make the use ML algorithms. Figure 7 depicts the schematic.

1. Map Function:

With these capacities, the information is perused and separated into a few parcels (sub-issues) that are then appointed to various processors. Every processor performs the map work on every segment (i.e., every processor performs its sub-issue). The map work gets a couple of (key1, value1) and returns rundown of (key2, value2) sets as a halfway outcome:

Map: (key1, value1) - - >(key2, value2)

Here (1) key1 and key2 are the keys with something very similar / various areas, and (2) value1 and value2 are the comparing esteems in certain spaces. From now on, these sets would be rearranged and arranged.

2. Reduce Function:

Every processor at that point executes the lessen work on both (1) a solitary key2 from this middle outcome and (2) the rundown of every quality related with key2 in the halfway outcome. All in all, every processor performs the diminish work on (key2, rundown of values2) to lessen — by consolidating, conglomerating, summing up, sifting, or changing—the rundown of qualities related with a given key2 (for all keys). Thus, a solitary (accumulated or summed up) value3 is returned:

Decrease: (key2, rundown of value2) - - >value3

Here (1) key2 is a key in certain areas, and (2) value2 and value3 are the comparing esteems in certain spaces. This MapReduce model use that clients just need to zero in on determining this map and diminish capacities without stressing over execution restraints for:

1. Parcelling the information
2. Across numerous machines planning and executing the program
3. Taking care of machine disappointments
4. Stray alleviation, or
5. Overseeing between machine correspondences.

To illustrate the findings, we used Statistical Programming (R). The R programming language is used to analyse data and assist researchers in quickly visualising the results.

2.3.2. Hadoop Distributed File System (HDFS)

Hadoop Distributed File System (HDFS) is a virtual evidence structure which is adaptable, runs on item equipment and gives high output admittance to application info. It is an information stockpiling segment of Hadoop. It stores its information blocks on top of the local record framework. It presents a solitary perspective on various actual plates or record frameworks. Information is circulated across the hubs; hub is an individual machine in a bunch and group is a gathering of hubs. It is intended for applications which need a compose once-read-many access. It doesn't permit change of information whenever it is composed. Hadoop has an expert/slave engineering. The Master of HDFS is known as Name hub and Slave is known as Data hub. The architecture of HDFS is illustrated in Figure 8.

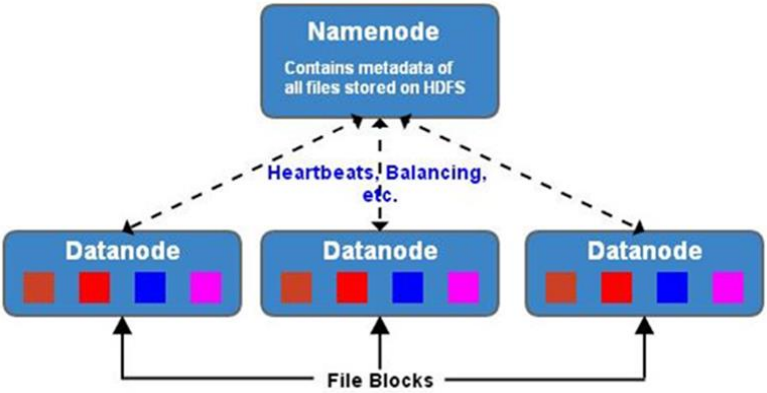


Figure 8. Architecture of HDFS.

Name Node

It is a deamon which runs on ace hub of hadoop bunch. There is just one name hub in a group. It contains metadata of the relative multitude of records put away on HDFS which is known as namespace of HDFS. It keeps two documents EditLog, record each change that happens to record framework metadata (exchange history) and FsImage, which stores whole namespace, planning of concludes to records and document structure properties. The FsImage and the EditLog are primary info designs of HDFS.

Data Node

It is a daemon which runs on slave machines of Hadoop group. There are number of information hubs in a bunch. It is answerable for serving perused/compose demand from the clients. It additionally executes block creation, erasure, and duplication upon assistance from the Name node. It additionally sends a Heartbeat message to the name hub intermittently about the squares it holds. Name node and Data node machines regularly run a GNU/Linux working framework (OS).

2.3.3. Essential Attributes of HDFS

The essential attributes of HDFS are presented in the following section.

1. Data Integrity:

At the point when a document is made in HDFS, it executes a checksum of every square of the record and stores this checksum in an altered secret

document. At the moment when a client recovers record element, it proves that the info it acquired matches the checksum clear up in the associated checksum document.

2. Robustness:

The vital target of HDFS is to store information reliably even inside the prospect of dissatisfactions. The three basic kinds of dissatisfactions are Name Node dissatisfactions, Data Node dissatisfactions and organization allotments.

3. Cluster Rebalancing:

The HDFS is viable with information re-adjusting that implies it will obviously move the information initial with one info hub into the next, if free space on information hub falls under a specific limit.

4. Accessibility:

It tends to be gotten to from applications from numerous points of view. Hadoop gives a Java API to applications to apply. A HTTP program can likewise be used to peruse the records of a HDFS case utilizing default web interface of hadoop.

5. Re-replication:

At the point when an information hub sends pulses to name hub and assuming any square is missing, name hub mark that block as dead. This dead square is re-recreated from the other information hub. Re-replication emerges when an information hub become inaccessible, a copy is undermined, a hard circle may fizzle, or the replication factor esteem is expanded.

2.3.4. Hive

Hive is data warehouse infrastructures tool that executes organized data in Hadoop and provides SQL like interface for processing/query the data. To analyse structured data, hive database technology is used to define databases & tables. This makes analysing and querying easy.

The main modules of Apache hive are:

- Hive Client
- Hive Services
- Processing and Resource Management
- Distributed Storage

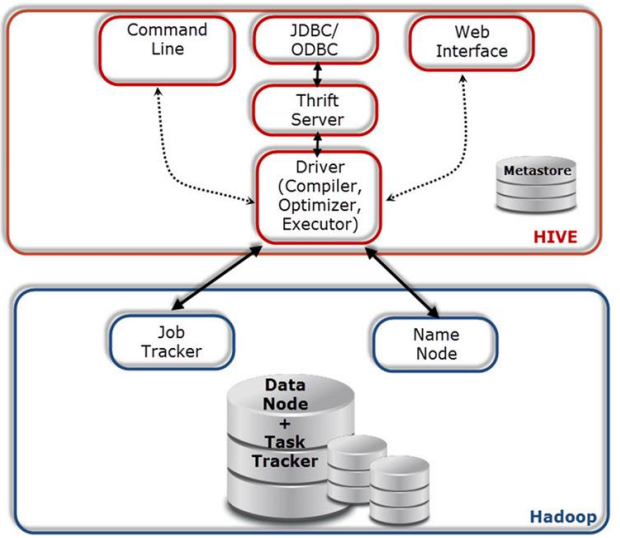


Figure 9. Big Data – Hive and Hadoop.

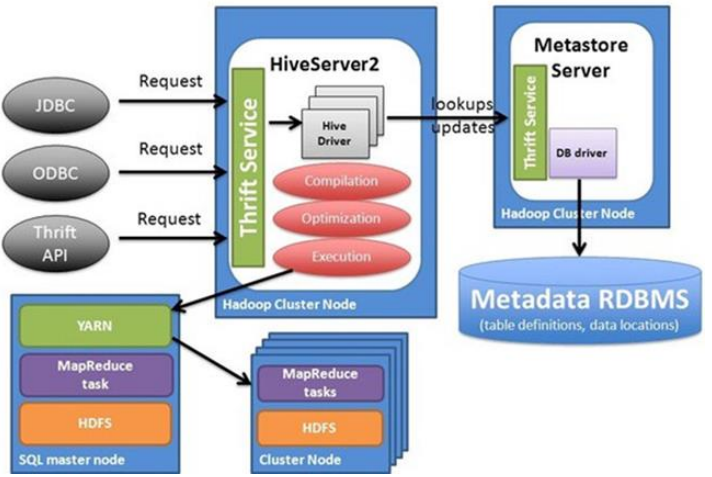


Figure 10. Architecture of a Hive.

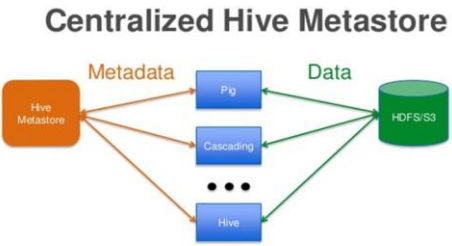


Figure 11. Hive Metastore.

Hive clients are Thrift, JDBC &ODBC client. Hive provides various services like the Hive server2, Beeline, Driver, Compiler, Optimizer, Execution Engine, Metastore, HCatalog and WebHCar. Hive's processing framework for performing the queries uses a software framework of MR. Hive uses Hadoop Distributed File System to process and store the data. Figures 9, 10 & 11 illustrate the working of Hive model.

2.4. Deep Learning Model

Deep Learning algorithms show promising results in the data analytics arena, along with distributed processing [6]. We have aimed for accuracy and performance improvement by implementing these techniques in our work. Kaur and Sharma (2019) have undergone a study on various classifiers for structure prediction. Among these, Deep Learning classifiers consist of a cascaded sequence of multiple layers with the capability to perform non-linear processing, automatic feature extraction, and transformation. Deep learning achieves a high level of abstracted features from low-level input features well as per Zhang and Kabuka (2019). Also, Deep Learning exhibits few representation learning properties [7]. These models accept raw data in vectored form like word embedding's, N-hot vectors, or image pixel values and the models know features from raw data automatically.

Feature extraction using handheld techniques degrade the performance of Deep learning models. Word embeddings-based motif vectors provide excellent help, which works best when implemented in a deep learning network [8]. Deep learning is one of the representation learning methods which can discover and learn hierarchical representations of data. Several deep learning architectures exist and this study uses a convolutional neural network that is trained based on motif embeddings.

After learning, Motif embeddings can be visualized as points in a multi-dimensional embedding space or one-dimension sequence logos [9]. We displayed that, from multi-dimensional space, evolutionary distant homologous sequences can be highlighted, and a comparison of homologous motifs can be visualized. Two metrics are used to compare the search results with current linear sequence motif-based methods [10]. Convolutional neural network departs from the motif extraction status quo, namely multiple sequence alignment algorithms, which require fine hand-tuning of motif search parameters. It automatically performs a search to produce results satisfying a given optimization criterion, such as maximum functional or structural similarity [11]. We investigated the generative ability of the convolutional neural network which resulted in development of computational methods that help in directed evolution of proteins towards altered or novel functions.

2.5. Proposed Methodology

Unstructured text data is an analogy with protein sequence data, and so NLP techniques are widely used in proteomics. Each letter in motif signifies an amino acid with which these NLP techniques convert amino acid into text form without losing motifs' position information. These techniques do not require traditional feature engineering with the advantage of giving good accuracy for any deep learning models. Inappropriate features lead to the poor performance of learning models, which degrades the deep learning network's accuracy.

2.6. Word Embedding's - Vectorisation

This vectorization technique's basic idea is that its neighboring word characterizes any term's meaning in a sentence. i.e., by its context. Skip Gram and CBOW are two variants of this external network, which uses a different type of input for generating the vectors.

- Skip-Gram: This is an unsupervised technique to predict the context word for a given target word. The results are the word embedding based vectors generated. By using the skip-gram model, the probability is maximized by adjusting network parameters.

- Continuous Bag of Words: CBOW is the same as Skip-gram. This is also an unsupervised technique to predict the target word based on context words.

In this study, we have represented our motifs using two way of vector representation. The first way is to create custom embeddings using input motifs [12]. Another way is to use inbuilt word embeddings motif2vec. motif2vec are embeddings, especially motifs that leverage higher-order, recurring, and statistically significant network connectivity patterns to transform the original graph to the motif graph.

3. AI Techniques for Motif Prediction

Proteins are a significant class of natural macro-molecules existing in the entire organic entities. All proteins have a novel direct arrangement of amino acids, likewise named a polypeptide. This amino corrosive grouping comprises data that manages the protein to crease up into a special form. Proteins overlap into one or more explicit spatial compliances in order to be able to play out their inherent capability. It is usually essential to find out the 3D architecture of proteins in order to appreciate their subatomic constituents. Among 20 amino acids, arginine, and lysine mediate sequence-specific DNA binding properties [13]. These basic residues can interact ionically with the negatively charged backbone phosphate groups or via hydrogen bonds with the bases. Figure 12 shows the common DNA binding motifs.

This research work includes NLP based feature representation technique, which involves the usage of prebuilt n-dimensional vectors from input protein structural data. These vectors are used as inputs to train the Deep Learning network to identify DNA binding motif [14].

Our experiments include all phases of data science, starting from data read, feature extraction, the analytical model build, and accuracy evaluation. A comparison study by Dencelin and Ramkumar (2017) on various learning approaches for protein secondary structure prediction is carried out. Protein data in FASTA format is being used throughout the research. The first phase of our study is the ingestion of protein data in FASTA format.

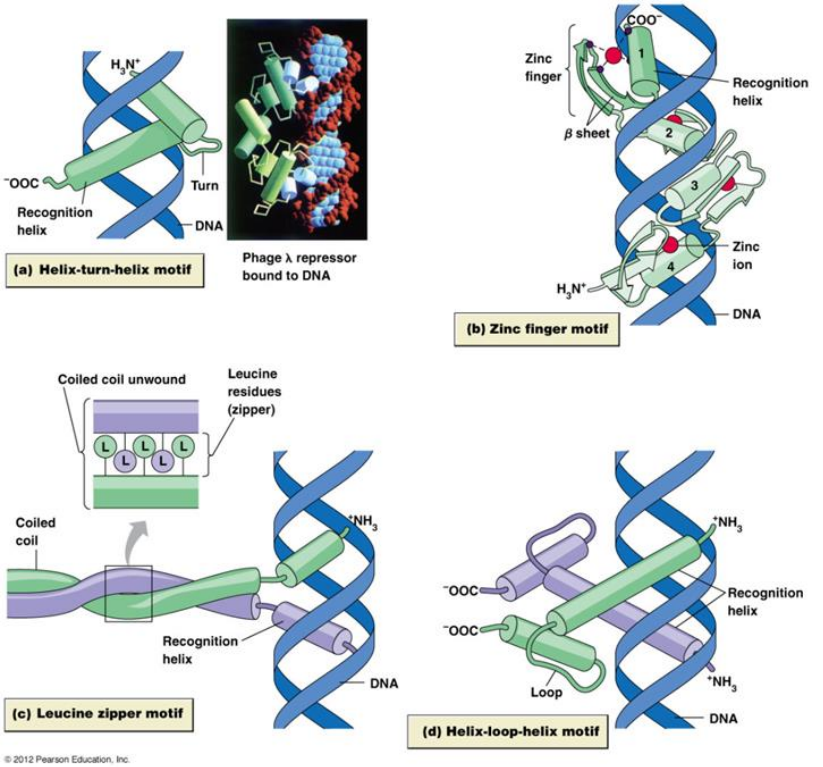


Figure 12. Various types of DNA binding motifs.

3.1. Input Dataset

Various motif dataset exists in the form of Regex or profiles HMM's. Publicly available motif databases include Pfam, eukaryotic linear motif database – ELM, ProCite, and ScanSite. FASTA format represents amino acids in single-letter codes for representing either nucleotide sequences or peptide sequences. The design begins with a single-line description, followed by lines of sequence data. For this experiment, we have considered the Helix-turn-helix DNA binding domain protein. FASTA files are pre-processed by removing white spaces and newlines. Invalid alphabets are also removed. A sample motif in FASTA format is shown in Table 1.

Table 1. G1BN16, A0A2D1GFQ0: Helix-turn-Helix's in FASTA Format

G1BN16	> tr | G1BN16 | G1BN16_9CAUD Helix-turn-helix DNA binding domain protein OS = Mycobacterium phage KLucky39 OX = 1034138 GN = 45 PE = 4 SV = 1 MTDQTHEAPPVPRYLSRDEVAKRLGMKSVRSLSGIELPPPDVEV GKHKGWLPETIDAWHA TRPGRGWWGGR
A0A2D1G FQ0	>tr|A0A2D1GFQ0|A0A2D1GFQ0_9CAUD Helix-turn-helix DNA binding domain protein OS=Mycobacterium phage Longacauda OX=2041540 GN=SEA_LONGACAUDA_44 PE=4 SV=1 MTDQTHEAPPVPRYLSRDEVAKRLGMKSVRSLSGIELPPPDVEV GKHKGWLPETIDAWHA TRPGRGWWGGR

3.2. Feature Extraction

Extracting the features is an excellent deal in Genome and Protein data, as the feature extractors are very specific in utility and function. The explicit feature engineering phase is not required, which is considered one of the critical pain points in Machine Learning, as Wei et al. (2019) pointed out. We used an unsupervised data-driven distributed representation technique to extract the features from input. Two variant of word embeddings has been used.

There are 3 phases of framework involved in motif2vec formation, as mentioned in Figure 13.

1. Network transformation: This phase is called the motif discovery phase, wherein we find instances of the motif(s). After receiving motif instances, a weighted motif adjacency matrix has been computed. This way, we transform the network to a motif graph, which encodes heterogeneity in network semantics and structure.
2. Sequence generation: Generate random walk sequence of original and motif(s) graph. Then aggregate and shuffle both sequences followed by feeding these sequences to the network
3. Embedding generation: The output of this phase is node embeddings. In this approach, we used the skip-gram network model.

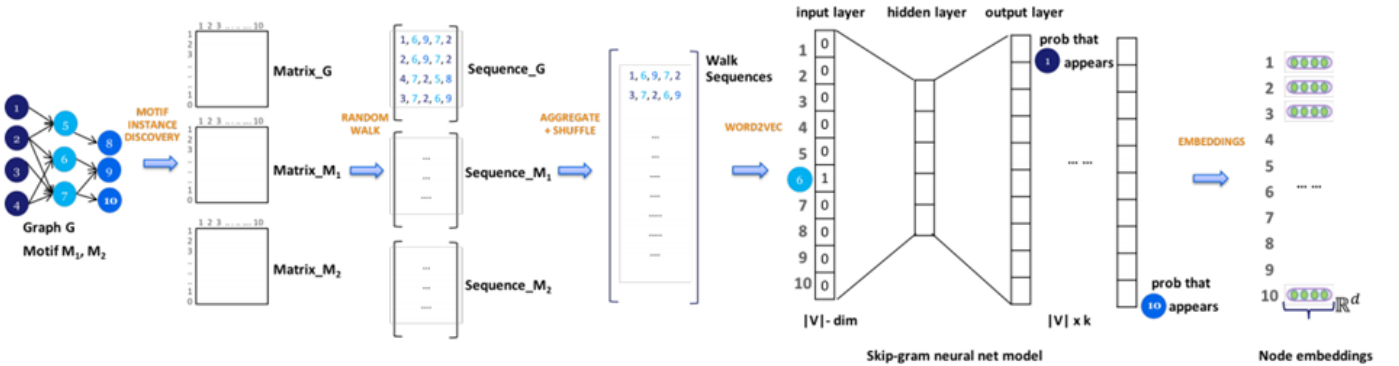


Figure 13. Motif2vec embeddings formation using deep learning network.

3.3. Predict Helix Turn Helix - AI Techniques

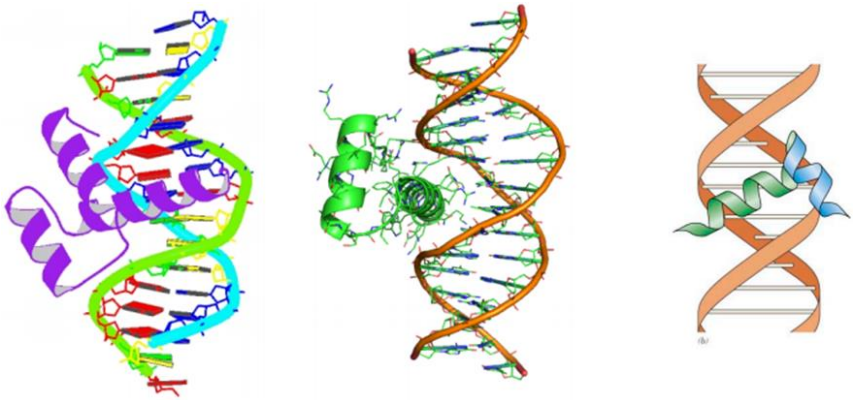


Figure 14. Structure of Helix turn Helix motif.

4. Deep Neural Network Implementation

Various sequence and structure characteristics are combined to predict motifs. Physio-chemical and structural properties commonly considered for prediction are.

- sequence conservation
- interface propensities

Features are finalized based on significance. The random forest variable importance technique is used to complete the most significant elements.

4.1. Convolutional Neural Network

CNN's core architecture is a network with rectification, followed by a Max-Pooling stage and a set of Fully Connected layers. The convolution stages have trainable motif detectors MD and rectification thresholds r, with weights W; the max-pooling stage has no trainable parameters. Essentially, a convolutional neural network model takes a single motif sequence M = (s1, s2, …, SL) and produces motif embedding vector M(V) = (m1, m2, .. . , mn),

where n is the number of filters used in the convolutional network. The motif embedding M(E) for sequence S is computed by a feed-forward flow of Convolutional Neural Network encoder, starting with convolution and ending in a fully connected neural network.

The architecture of convolutional neural networks differs based on neural network type, a number of motif detectors, and input dataset.

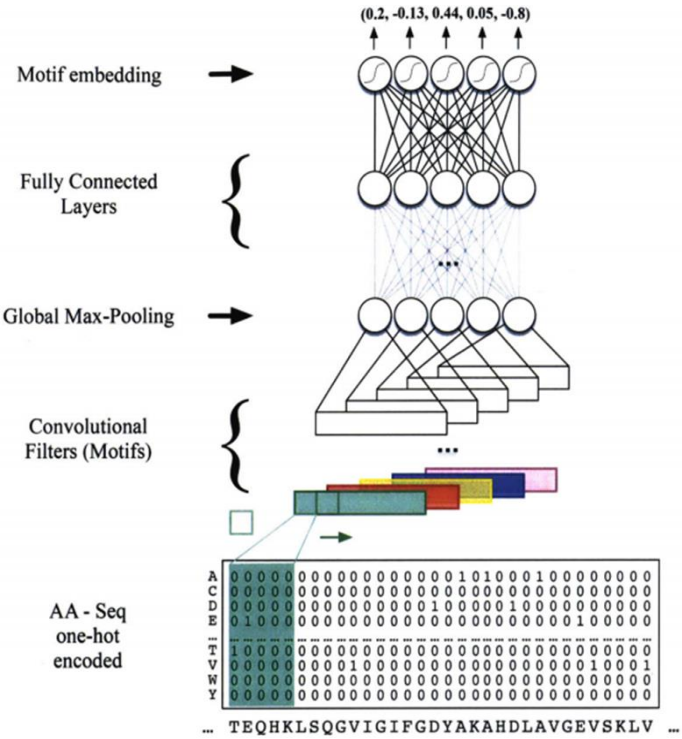


Figure 15. CNN architecture which encodes protein sequence to motif encoding and motif identification.

The architecture diagram in Figure 15 provides the pipeline of the motifs identification process. It involves motif embeddings followed by a fully connected hidden layer, max-pooling layer, and convolutional layer. The entire framework is implemented in SPARK for improved performance. The output is the class labels with motifs tagged.

For the success of any deep learning network, proper tuning of network parameters is required. Among the various network parameters, the learning rate and several filters play a major role in calculating the accuracy, and these

are called hyper-parameters. To optimize the hyper-parameters of CNN, we mainly focused on the below,

No. of motifs and its length: The count of the motif and its size vary significantly based on the motifs' characteristics. Figure 16 shows the length distribution of motifs that exists in PROSITE DB. A peak occurred at 15, but with motifs, more than 300 amino acids exist. Factors influence are:

a) Motifs and it's biological interpretation
b) More than one occurrence in the same sequence
c) Input dataset and its sequence diversity
d) Motif hierarchy

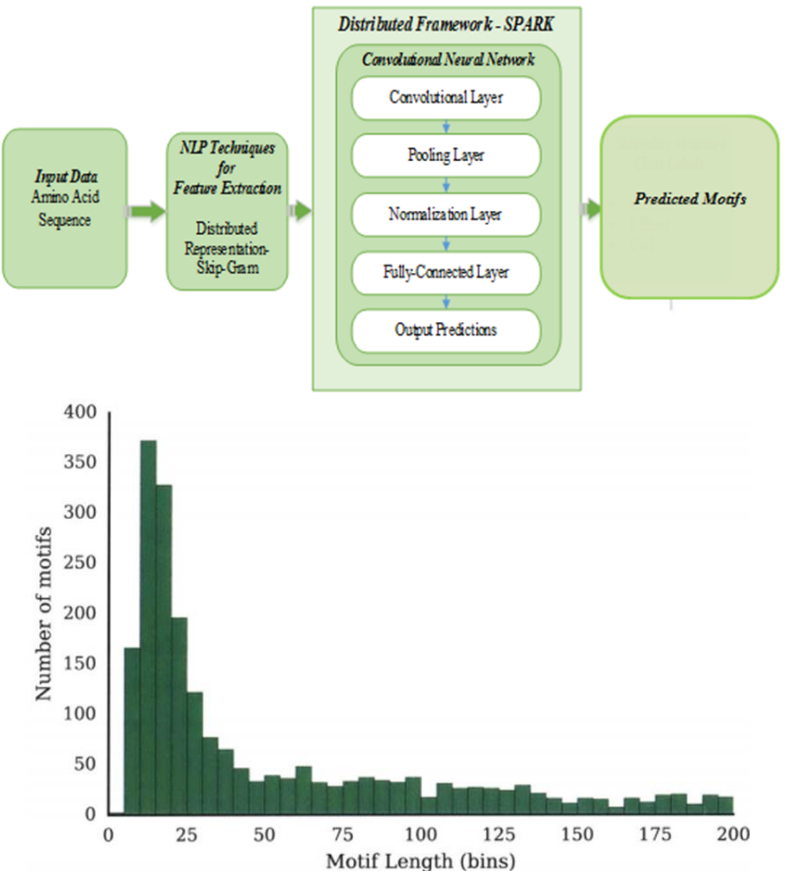


Figure 16. Length distribution of motif from the UniProtKB database.

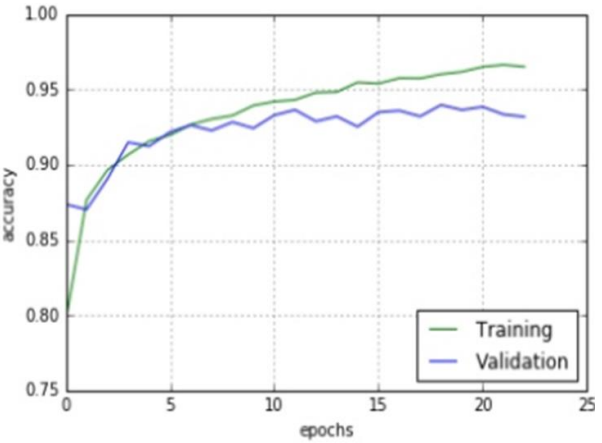


Figure 17. High-level architecture of motif identification framework.

No. of convolutional and fully connected layers: This mainly depends on the size of the motif dataset as it depicts the order of magnitude. Increasing the fully connected layers provides a way to include more complex motifs and to identify complex relationships.

Figure 17 shows high-level architecture with various layers of Convolutional Neural Network.

5. Experimental Results

Python was used for computation to implement Machine Learning and Deep Learning algorithms. We have used a confusion matrix to evaluate the accuracy of the convolutional neural network classifier. Various network parameters are modified to improve the accuracy of identified motifs. We have used different learning rates starting from smaller rates (0.01) and was gradually increased with varied training sets and the number of network layers. The model took time to converge for the lower learning rates. After that, we have increased the learning rate and found that the loss becomes high. Again, we have gradually decreased the learning rate after each epoch with different training sets, thereby facilitate faster training. The learning rates are modified based on two schedules:

- Drop Based
- Time-Based

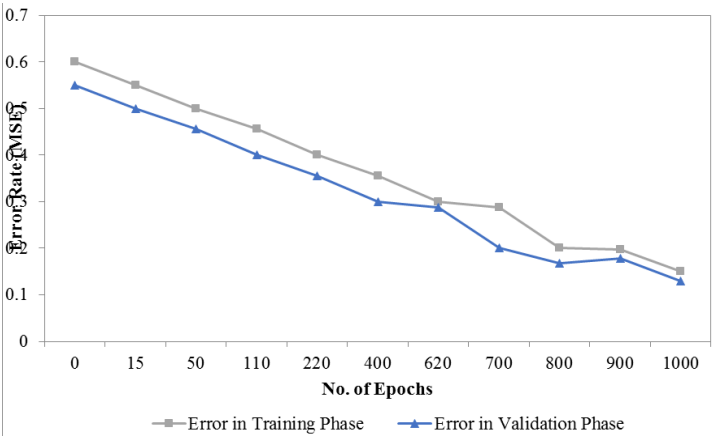


Figure 18. Varied MSE values and Epochs for accuracy measurement.

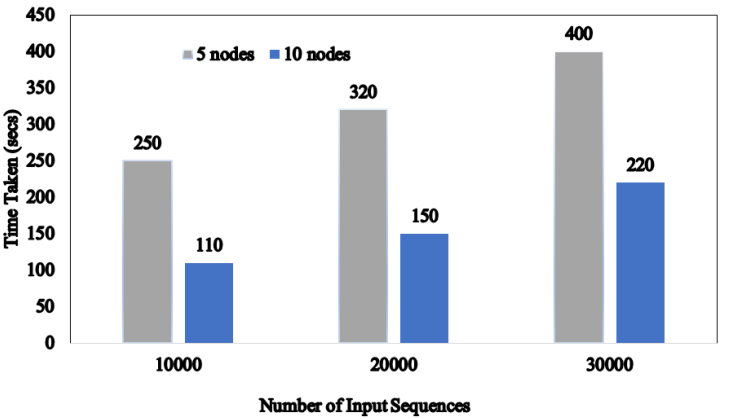


Figure 19. Performance of cluster with a varied number of input sequences.

Length of motifs and the number of motifs are varied and evaluated for better accuracy. Structural and sequential conservation is tested by having an order of 10-20 and 40 – 100 amino acids. Also, different values of epochs are tested and evaluated for the accuracy of the convolutional neural network. We mainly used epoch ranges from 50-500, and the corresponding accuracy measures are captured in Figure 18. We noticed that the MSE values are pretty good with increased epochs.

To improve performance, we used the SPARK framework. The account can be increased by adding more nodes to the cluster as in Figure 19.

5.1. Confusion Matrix Results

Confusion matrix entries for various classes are mentioned in Table 2.

Table 2. Confusion matrix entries of helix turn helix motifs types

The proportion of predicted helix-turn-helix				
Actual	*Predicted*			
		Di-Helical	Tri-Helical	Tetra-Helical
	Di-Helical	75.80%	10.67%	8.77%
	Tri-Helical	15.57%	78.43%	3.68%
	Tetra-Helical	14.76%	2.97%	82.74%

Table 3. Abinitio approach vs. NLP based approach

Approaches and Accuracy comparison		
Deep Learning Model	Feature Extraction	Accuracy
Deep Neural Network	Physio-chemical Evolutionary: sequence conservation	82.5%
Convolutional Neural Network	NLP based Feature Representation with prebuilt vectors (motif2vec).	83.4%
Convolutional Neural Network using ensemble word embeddings	Motif2vec + inbuilt vectors	84.7%

We have evaluated a few approaches for predicting DNA binding motifs using Deep learning architectures. As it is shown in Table 3, the Abinitio methodology that has been proposed by Spencer, Eickholt, and Cheng achieved (80.7%) Q3 accuracy. This approach provides an extra input profile extraction phase, which extracts the Evolutionary sequence conservation. In this approach, different windows sizes with different combinations of the input features are considered to train the Deep Learning Network. RBIs are used here. Table 3 shows that our approach achieved 84.7% accuracy on the test set sequences using the UniProtKB protein dataset, which contains annotations of the secondary structures along with the primary series. The proposed model suggests that the feature extraction using ensemble NLP technique followed by Convolutional Neural Network classifier is well suited to identify DNA binding motifs, thereby attaining better accuracy.

Conclusion

The research outcome of this approach for the prediction of DNA binding motif with CNN provides an accuracy of 83.9%, 84.31% sensitivity, 86% specificity. NLP, together with deep learning, is more accurate than other Machine Learning techniques. NLP improves accuracy and speed, which is required for the efficient proteomic study. The application of in-depth knowledge in bioinformatics provides performance and precision by investigating new algorithmic methods for identifying the functions, with more parallelized methods to improve computational performance.

References

[1] Rithvik, M., and Rao, G. N. (2015). A Comparative Study of Methodologies of Protein Secondary Structure. In *Computational Intelligence Techniques for Comparative Genomics* (pp. 37-45). Springer, Singapore.

[2] Li, Z. C., Zhou, X. B., Lin, Y. R., and Zou, X. Y. (2008). Prediction of protein structural class by coupling an improved genetic algorithm and support vector machine. *Amino acids*, *35*(3), 581-590.

[3] Xavier, L. D., and Thirunavukarasu, R. (2017). A distributed tree-based ensemble learning approach for efficient structure prediction of protein. *Training*, *227*.

[4] Bouziane, H., Messabih, B., and Chouarfia, A. (2015). Effect of simple ensemble methods on protein secondary structure prediction. *Soft Computing 19* (6), 1663-1678.

[5] Yaseen, A., and Li, Y. (2014). Context-based features enhance protein secondary structure prediction accuracy. *Journal of chemical information and modeling*, *54*(3), 992-1002.

[6] Collobert, R., and Weston, J. (2008). A unified architecture for natural language processing: Deep neural networks with multitask learning. In *Proceedings of the 25th international conference on Machine learning* (pp. 160-167).

[7] Asgari, E., and Mofrad, M. R. (2015). Continuously distributed representation of biological sequences for deep proteomics and genomics. *PloS one*, *10*(11), e0141287

[8] Le, Q., and Mikolov, T. (2014). Distributed representations of sentences and documents. In *International conference on machine learning* (pp. 1188-1196).

[9] Hu, B., Lu, Z., Li, H., and Chen, Q. (2014). Convolutional neural network architectures for matching natural language sentences. *Advances in neural information processing systems*, *27*, 2042-2050.

[10] Xu, J., Xiang, L., Hang, R., and Wu, J. (2014, April). Stacked Sparse Autoencoder (SSAE) based framework for nuclei patch classification on breast cancer histopathology. In *2014 IEEE 11th International Symposium on Biomedical Imaging (ISBI)* (pp. 999-1002). IEEE.

[11] Wang, Y. B., Yu, Z. H., Li, X., Jiang, T. H., Chen, X., Zhou, X., and Wang, L. (2017). Predicting protein-protein interactions from protein sequences by a stacked sparse autoencoder deep neural network. *Molecular BioSystems*, *13*(7), 1336-1344).

[12] Szczerba, M., Wiewiórka, M. S., Okoniewski, M. J., and Rybiński, H. (2016). Scalable Cloud-Based Data Analysis Software Systems for Big Data from Next-Generation Sequencing. In *Big Data Analysis: New Algorithms for a New Society* (pp. 263-283). Springer, Cham.

[13] Zhang, C., De Sterck, H., Aboulnaga, A., Djambazian, H., and Sladek, R. (2009, June). Case study of scientific data processing on a cloud using Hadoop. In *International Symposium on High-Performance Computing Systems and Applications* (pp. 400-415). Springer, Berlin, Heidelberg.

[14] Spencer, M., Eickholt, J., and Cheng, J. (2014). An in-depth learning network approach to ab initio protein secondary structure prediction. *IEEE/ACM transactions on computational biology and bioinformatics*, *12*(1), 103-112.

[15] G. Sivagamasundari and Dr. Latha Parthiban, "Clustering Protein Super Secondary Structure with Artificial Neural Network Based Self Organizing Map Machine Learning Approach Using In-Memory Computing Environment" *Journal of Computational and Theoretical Nano-science* Vol. 17, 1–3, 2020.

[16] D. Shine Babu, Dr. Latha Parthiban, G. Sivagamasundari, "Motif Structure Prediction in Distributed Framework using Machine Learning Algorithms" *International Journal of Recent Technology and Engineering* (IJRTE) ISSN: 2277-3878, Volume-7, Issue-5S4, February 2019.

Chapter 6

Quantum Behaved Translation Invariant Feature Extraction for Chromosome Classification

S. Nithinsha[1,*] and D. Vinod[2,†]

[1]Department of CSE, Saveetha School of Engineering, SIMATS, Chennai, India
[2]Department of Cloud and Edge Analytics, Institute of Computer Science and Engineering, Saveetha School of Engineering, Chennai, India

Abstract

Automation of chromosome classification is needed for genetic disease diagnosis and cancer pathology research. Several approaches encompassing Statistics, Neural Networks, and Artificial Intelligence have been attempted for the efficient analysis and classification of chromosomes. Each has reported its share of merits and demerits While Karyotyped images of chromosomes are less susceptible to scaling transformations, experimental and handling errors during the analysis may subject the images to translation and rotation transformations. The first level feature vectors extracted from preprocessed Karyotyped images of chromosomes are rendered invariant to the translation by transforming the coordinate frame to the centroid of the image. Further, Ring projection of the images not only renders the feature vectors invariant to the rotation but also ensures dimensionality reduction by way of transforming the two-dimensional images to one-dimensional signals. The subsequent application of Daubechies wavelet transform to the signals decomposes the same into sub-patterns suitable for use by the classifier. In

* Corresponding Author's Email: ayshaas2017@gmail.com.
† Corresponding Author's Email: dvinopaul@gmail.com.

In: Frontiers in Quantum Computing
Editor: R. Anandan
ISBN: 978-1-68507-816-4
© 2022 Nova Science Publishers, Inc.

this work, a simple minimum distance classifier has been employed. The feature extractor results in a classifier performance of 100% recognition rate.

Keywords: pattern recognition, chromosome classification, wavelet transform, neural networks

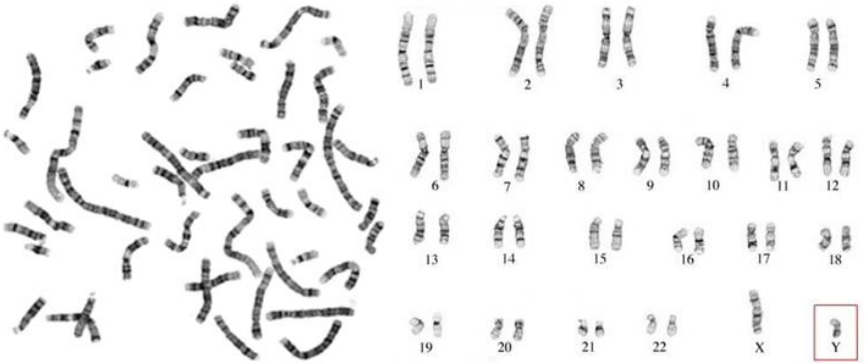
1. Introduction

Pattern Recognition Systems (PR) have found wide applications in signal processing, image analysis, communication systems, biomedical imaging, and control systems, to quote a few [1]. The heart of any PR system is feature extraction. Features serve to identify one class of patterns from another, thereby aiding better pattern recognition and/or classification. Also, the objective of any feature extractor is to characterize the object and reduce the dimensionality of the pattern space by mapping the original patterns to the feature space. Again, feature extractors need to be invariant to image transformations such as rotation, translation, and scaling which are bound to occur due to experimental constraints or handling errors. Thus the hallmark of any feature extractor is

I. characterization of objects or patterns
II. reduced dimensionality of feature space
III. invariance to pattern transformations

In the field of clinical cytogenetics, automated chromosome classification has been an important PR problem for decades [2]. Human chromosome analysis is an essential task in prenatal screening, genetic disease diagnosis, and cancer pathology research. The cytogenetic analysis begins by breaking down the nucleus of the human cell into a groggy mass of chromosomes which are then stained with a dye (Giemsa), to be observed under a microscope. These Giemsa banded or G banded chromosomes are photographed creating a chromosome spread image as shown in Figure 1(a). This image of the clustered chromosome undergoes a process called Karyotyping which is a standard form of display of the individual chromosomes in the cluster. The display as shown in Figure 1(b) configures the 22 pairs of chromosomes and the sex chromosomes X, Y in humans, according to the order of the length – longest in chromosome 1 to the shortest in chromosome 22. Chromosome 23 is the pair of sex chromosomes. Karyotyped images are used to find whether

the patient's chromosomes are normal or abnormal, the latter suggestive of some genetic disorders or complications. Karyotyped images of chromosomes are routinely used in clinical and cancer cytogenetic labs. The Karyotyping of chromosomes involves segmentation, enhancement, orientation, measurement, and classification, thereby calling for modern image processing techniques. However, the problem of classification which is the final decision in the process has been an outstanding PR problem. Commonly, the classification is done by visual inspection by the cytogeneticist and is laborious and subject to human error.



(a) Chromosome spread image (b) Karyotyped chromosome images

Figure 1. Chromosome images before and after Karyotyping.

Several investigators have proposed methods to tackle the problem of automating chromosome classification. Detcheva [3, 4] proposed an object-oriented intelligent approach and a heuristic knowledge-based approach to the problem. Jennings and Graham [5] undertook a preliminary investigation on the application of two different Neural Network (NN) architectures on a natural representation of the chromosome banding patterns and concluded that the Multilayer Perceptron (MLP) showed better classification performance than Kohonen's Self-organizing Feature Map. Lerner et al., [6] also investigated the application of MLP for automated chromosome classification. Lerner [7] showed a classification performance of 83.6% with his technique. Mousavi et al., [8] investigated multicolor images of metaphase chromosomes using centromere segmentation. Cho [9] applied Backpropagation Neural Networks to chromosome classification and reported minimum classification error for certain configurations of the network. However, the investigation reported problems such as long training sessions, manual determination of the

centromere locations for telocentric chromosomes, and trial and error experiments to obtain the architecture of the network. In recent years, Wu and Castleman [10] made use of wavelet-based band descriptors for the automation of chromosome classification.

In this investigation, we propose a wavelet-based feature extractor invariant to rotation and translation transformations, for the efficient classification of chromosomes. In the automation of chromosome classification, there is lesser room for scaling transformation since the Karyotyped images are all set to the same scale. However, during the pattern recognition of the test images, there is greater scope for the image transformations of translation and rotation to occur calling for the invariance of the feature extractor to these two transformations. Ring Projection [11, 12] has been employed to explicitly tackle the problem of rotation invariance. Wavelet coefficients of the pre-processed chromosome images serve to extract the feature vector sets. Figure 2 illustrates the overall approach of chromosome classifier. The feature extractor when utilized for the classification of normal chromosomes reported a 100% classifier performance.

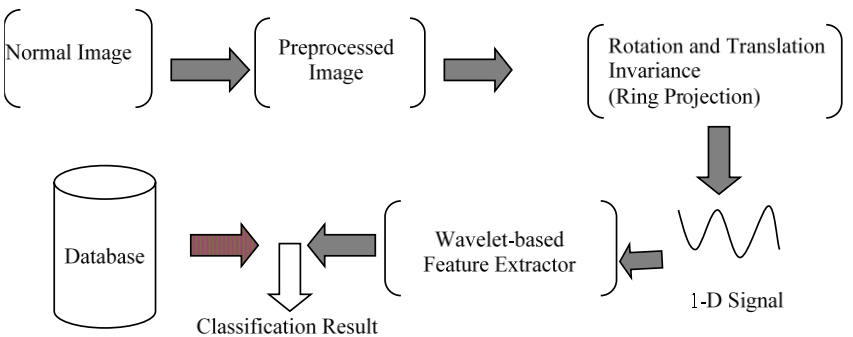


Figure 2. Block diagram of the Wavelet-based translation and rotation invariant feature extractor for image classification.

Section 2 introduces the rotation and translation invariance of the feature vectors and Section 3 microarray data along with the structure of the automated chromosome classifier. Section 4 details denoising with adaptive thresholding and in Section 5 conclusions are presented.

2. Rotation and Translation Invariant Feature Extraction

Most PR-based image processing applications are subject to transformations such as rotation (R), scaling (S) and translation (T) due to handling/ experimental errors and the feature extractor must display the property of invariance to these transformations for efficient recognition/classification of patterns. Several methods based on moment functions [13] have been proposed. Rajasekaran and Pai [14, 15, 16] modified Hu's moments to exhibit invariance to both S and R transformations. The moment invariant-based feature extractor had been applied for chromosome classification with Kohonen's SOFM as a classifier with reasonably good performance [17].

In this investigation, considering the recognition/classification of the Karyotyped images of chromosomes which are set to the same scale, it was only essential that the feature extractor remains invariant to R and T transformations [18]. To achieve T transformation, it was merely required that the centroid of the complemented image be computed and the reference coordinate framework translated to the centroid as the origin. Figure 3 illustrates the same. For invariance to R transformation, ring projection has been employed [19]. The discretization of the grayscale image I(x,y) of a chromosome into its binary equivalent is given by

$$I(x,y) = \begin{cases} 1 & if\ (x,y) \in D \\ 0 & otherwise \end{cases} \tag{1}$$

where *D* is the domain corresponding to the white region of the image as shown in Figure 3. From equation (1) it is evident that the multivariate function *I(x,y)* is a two-dimensional density function of mass distribution on the plane and also follows a uniform distribution. In other words, the mass is homogeneously distributed over the region *D*. Let $m(x_0, y_0)$ be the centroid as shown in Figure 3. Now we obtain

$$M = \max_{p \in D} |p(x,y) - m(x_0, y_0)| \tag{2}$$

where $|p(x,y) - m(x_0, y_0)|$ is the Euclidean distance between two points *p(x,y)* and the centroid $m(x_0, y_0)$ on the plane. Transforming the Cartesian coordinate frame to polar coordinates using the relation,

$$x = r\cos\theta\,, \quad y = r\sin\theta \tag{3}$$

we obtain,

$$I(x, y) = I(r\cos\theta, r\sin\theta) \tag{4}$$

where $r \in [0, \infty)$, $\theta \in (0, 2\pi]$.

Now for any $r \in [0, M]$, $f(r)$ which is the total mass as distributed along with the circular rings as shown in Figure 4 is given by

$$f(r) = \int_0^{2\pi} I(r\cos\theta, r\sin\theta)d\theta \tag{5}$$

in the continuous case and given by

$$f(r) = \sum_{i=0}^{M} I(r\cos\theta_i, r\sin\theta_i) \tag{6}$$

in the discrete case. $f(r)$ is termed as the ring projection of the planar mass distribution and as is evident is a one-dimensional representation of the two-dimensional image transformed through ring projection [20, 21]. Figure 4 illustrates the ring framework for a chromosome and its rotated version and Figure 6 the ring projection of the same.

Thus the transformation of the coordinate frame to the centroid of the pattern coupled with the ring projection ensures invariance to both T and S transformations besides transformation of the two-dimensional image patterns to one-dimensional signals.

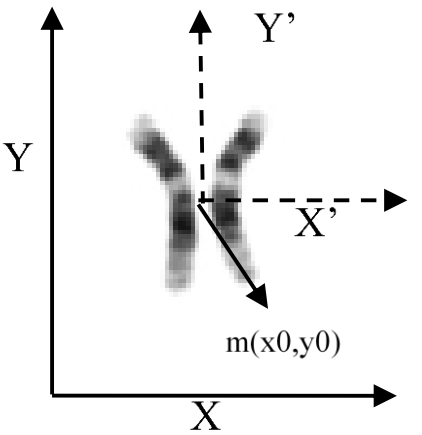


Figure 3. Translation of reference.

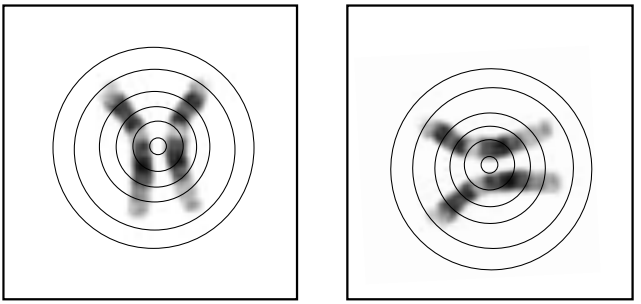


Figure 4. Ring frames on an image and its rotated Coordinates to the centroid m(x0, y0) version.

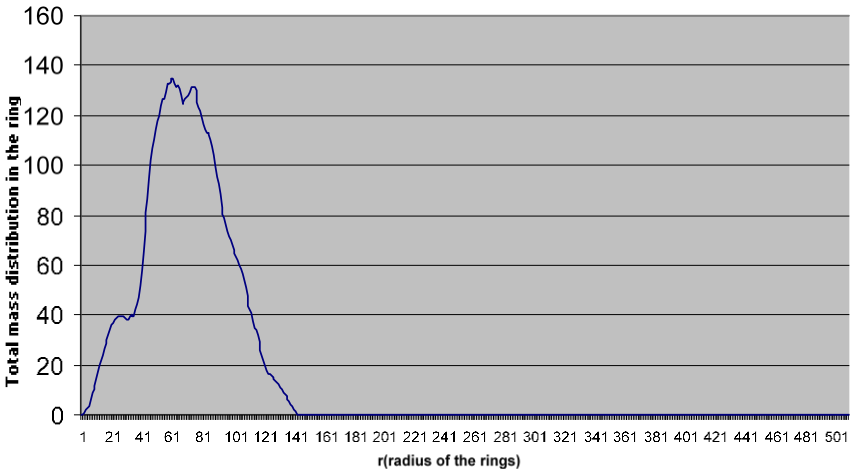


Figure 5. Ring projection of the chromosome(s) shown in Figure 4.

2.1. Wavelet Orthonormal Decomposition into Subpatterns

In this work, Daubechies Wavelet transform of level 2 (Daub2) has been employed to transform the one dimensional signal of the chromosome image patterns obtained through ring projection, into a set of subpatterns. It is these wavelet decomposed sub-patterns that will form the feature vector set for final classification. The approximate coefficients (*approx2*, *approx3*, *approx4*) and detail coefficients (*det2*, *det3 det4*) of Daub2 up to a level of 4 have been extracted. Figure 6 illustrates the coefficients of the one-dimensional signal of the chromosome illustrated in Figure 5. However, for chromosome

classification application it was observed that the approximate coefficients (*approx2*, *approx3*, *approx4*) and one among the detail coefficients, viz., *det2* was more than sufficient to ensure good classification results. Table 1 shows the feature vector of *approx4* coefficients for the chromosomes listed and their transformations. Observe the proximity among the coefficient values. Also, the tolerance limits for the correct classification of chromosomes must be arrived at experimentally to ensure a better recognition rate. The tolerance limit is the least upper bound of the difference in distances between a test chromosome and the nominal chromosome sets in the database. Table 2 illustrates the tolerance limits arrived at experimentally for the chromosome sets ensuring their correct classification.

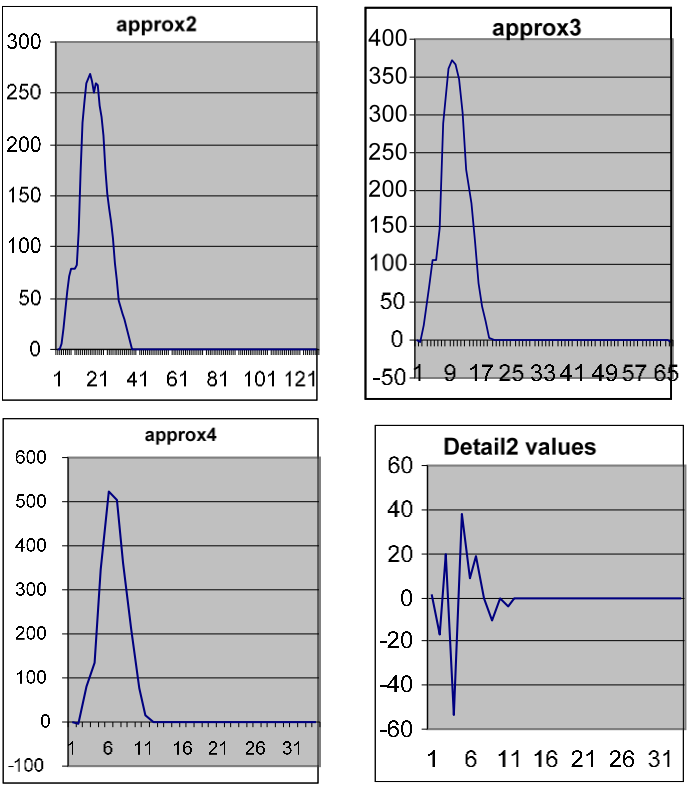


Figure 6. *Approx2*, *Approx3*, *Approx4*, and *det2* coefficients of the one-dimensional signal of Figure 5.

Table 1. Approx4 coefficient values for sample chromosomes – nominal and rotated

	Approximate coefficients (*Approx4*)					
Chromosome1 (Nominal)	27.70	26.67	137.66	306.81	300.95	285.26
	302.30	260.97	224.18	216.08	136.55	40.01
Chromosome1 Rotated 45°	27.11	25.65	139.24	307.73	298.38	285.51
	303.46	261.41	223.68	215.87	137.02	38.91
Chromosome1 Rotated 315 °	26.29	25.46	141.80	304.53	298.42	286.96
	307.26	261.65	226.58	217.45	136.09	38.95
Chromosome18 (Nominal)	-0.69	-11.71	99.95	327.72	189.63	11.01
	0.00	0.00	0.00	0.00	0.00	0.00
Chromosome18 Rotated 45°	-0.70	-12.39	102.19	329.69	191.59	11.49
	0.00	0.00	0.00	0.00	0.00	0.00
Chromosome18 Rotated 315°	-0.67	-12.24	101.24	330.41	189.71	12.19
	0.00	0.00	0.00	0.00	0.00	0.00

```
Procedure MIN_DIST_CLASSIFY(chro, limit, newchro)
/* chro[i,j] is a vector representing the jth coefficient of the ith chromosome where
1≤j≤4 and 1≤i≤24. The coefficients considered are approx2, approx3, approx4 and
det2.
newchro[j] is the test chromosome to be classified where 1≤j≤4 indicates its
approximate and detail coefficients.
diffchro[i,j] indicate the sum of the absolute vector difference between the jth
coefficients of the ith chromosome and the test chromosome*/
begin
for i=1 to 24
for j=1 to 4
diffchro(i, j)= ∑ |chro[i,j]-newchro[j]|;
for j=1 to 4
obtain min[j] as index i for which diffchro[i, j] is minimum;
If (min[1]=min[2]=min[3]=min[4])
   {choose k to be one of min[j], 1≤j≤4;
        Let limit[i,j] , 1≤j≤4 and 1≤i≤24, be the computed tolerance
        limits obtained through experimental analysis for the ith
        chromosome and jth coefficient;
        If(diffchro[k,j])<limit[k,j] for all 1≤j≤4
        print (“Test chromosome identified is chromosome k”);
        else print(“Test chromosome is unknown”);
}
print(“Test chromosome is unknown”);
end
```

Figure 7. Algorithm for minimum distance classifier.

Table 2. Tolerance limits for the chromosomes feature vector sets ensuring correct classification

Chromosome	Limits				Chromosome	Limits			
	Approx2	*Approx3*	*Approx 4*	*Detail 2*		*Approx2*	*Approx3*	*Approx 4*	*Detail 2*
1	154.99	107.33	55.70	75.65	13	99.27	73.04	50.24	41.52
2	167.92	117.73	66.35	71.27	14	64.89	45.14	24.14	39.34
3	145.10	100.14	45.54	67.99	15	46.98	29.04	18.03	27.87
4	125.49	87.53	57.77	55.09	16	39.89	24.28	12.69	28.23
5	89.77	52.85	26.46	55.19	17	41.33	25.55	17.99	21.05
6	109.01	68.84	47.43	51.38	18	36.64	22.76	11.97	25.68
7	92.09	55.63	25.05	47.65	19	27.69	20.31	8.35	17.16
8	96.26	58.17	33.67	42.72	20	49.52	31.73	14.04	22.63
9	85.60	63.59	48.45	40.54	21	34.20	21.85	13.10	19.48
10	68.06	43.68	26.54	30.36	22	31.35	17.53	10.61	17.43
11	88.82	59.92	22.91	42.02	X	47.23	34.11	24.83	26.33
12	73.78	51.23	30.41	39.95	Y	23.78	18.39	10.78	13.73

2.2. Performance Analysis

To analyze the performance of the wavelet-based feature extractor on the classifier, the following experiments were undertaken. In the first experiment a test set of a specific chromosome rotated by angle θ, $0 < \theta < 2\pi$and with random translations was tested for its recognition. Figure 9 shows the minimum distance computed for a specific test chromosome (chromosome 1) when presented with various rotations and translations against a nominal pattern of chromosome 1 in the database. With the experimented tolerance limit of 155 chosen for the correct identification of chromosome 1, it was observed that the minimum distance of all the test sets lay within the tolerance limit thereby ensuring 100% recognition of the test set. In the next experiment, the impact of the tolerance limit value on the correct classification of the chromosomes was studied. As mentioned earlier for the tolerance limit value of 155, for the same test set of transformed chromosome 1 patterns, the classification percentage was 100%. However, a decrement in the tolerance limit increased the misclassification of the test chromosomes. Figure 10 illustrates the association between the tolerance limit and the percentage of correct classification for a specific test set.

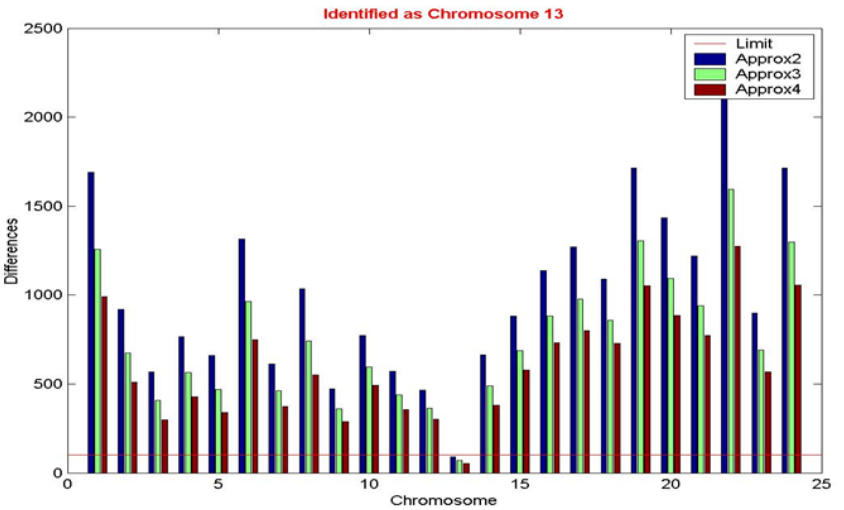


Figure 8. Recognition of a test chromosome as chromosome 13.

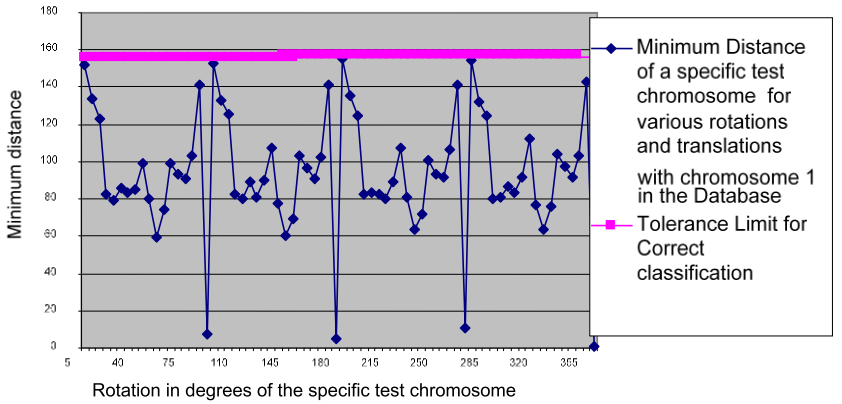


Figure 9. Recognition of chromosome 1 for translation and rotation transformations.

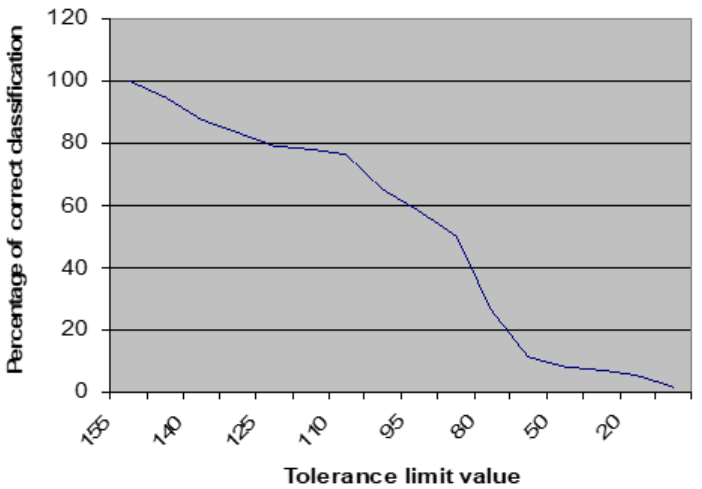


Figure 10. Impact of tolerance limits on the recognition of chromosomes.

(iii) The tolerance limits work as least upper bounds for the correct classification of chromosomes.

(iv) The application has been tested over G-banded chromosome photographic images.

Further investigations are in progress to employ a Neural Network based classifier to dispense with the tolerance limits that the minimum distance classifier requires.

3. Microarray Data

Microarray methodology has made it conceivable to consider the articulation stages of thousands of genes at the same time. In this work, genuine information gathered from the clinics of various patients has been utilized. The patient's information has been contrasted and the expert's information has been demonstrated by heat map arrangement. An appropriate investigation of gene expression information focuses on cancer recognizable proof and arrangement, aids in legitimate treatment choice.

Table 3. Data Set Input

ID_REF	IDENTIFIER	GSM627133	GSM627216	GSM627134	GSM27151	GSM27115
231224_at	PRKAG2	35.8955	30.7351	32.9837	32.7659	37.144
240882_at	R85522	11.6371	10.5217	10.8537	11.8015	11.9516
1561849_at	PKD1L2	8.45343	8.46326	7.39276	7.33646	7.37014
1565746_at	LOC100132815	10.8116	11.1259	9.72156	8.65285	12.1602
1560853_at	ZNF826P	16.0361	16.6499	14.5228	15.519	17.4264
230660_at	SERTAD4	12.4425	14.2759	12.7199	13.7253	12.7716
229708_at	TOR4A	13.3051	13.8157	12.2562	19.5204	14.6117
244781_at	R37682	8.12955	10.2744	8.59731	10.2062	9.19906
1554187_at	LOC554206	16.9199	15.6306	18.8628	17.3767	14.8512
230021_at	TICRR	25.4832	32.9972	27.267	26.6506	25.6564
238226_at	TMEM255B	29.5791	26.6391	27.1048	25.821	27.9684

Gene expression analysis using microarray technology has yielded useful cancer detection results. The suggested system collects several sources of data and creates associated storage, clusters the data, and classifies it. DNA Microarrays and heat maps were utilized to look at a smaller number of genes rather than the full genome. Cancer prediction is simple using the proposed methodology, as long as the heat map reveals the afflicted levels. An appropriate examination of gene expression data could provide sufficient information to identify cancer and also to classify cancer. This helps in the proper selection of appropriate cancer treatment.

Figure 11. Heat Map Gene Expression Data Prediction.

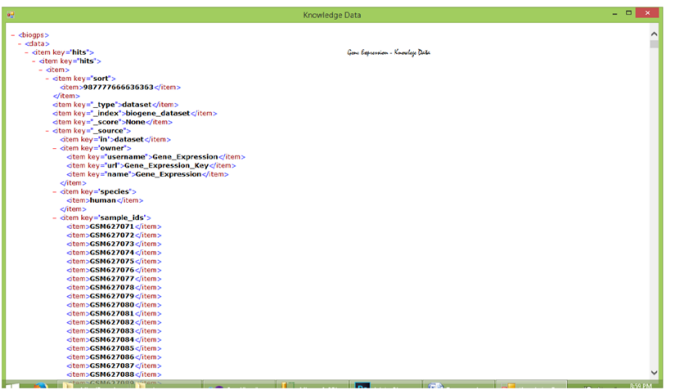


Figure 12. Knowledge Data.

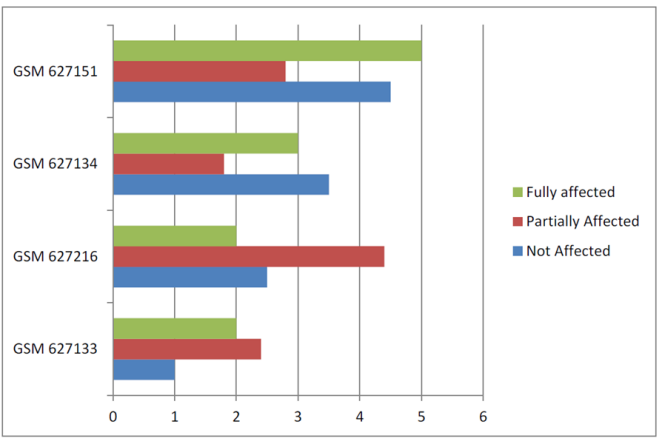


Figure 13. Heat Map Showing the Affected Levels.

3.1. Cluster Performance Analysis

The performance of the Self Organizing Map (SOM) relies on network parameters and the features used. The evaluation of the various accuracy measures using a different set of network parameters and the results are plotted as shown in successive sections.

For predicting super secondary motifs, Back Propagation (BP) neural secondary structure prediction is used. Studies on predicting the secondary structures using ANN-based technique namely multilayer perceptron were also done. Features are selected from the protein sequence which is amino acids. Then, the parameters for the SOM clustering are identified which is implemented in an in-memory computing SPARK environment.

3.2. Protein Dataset

Dataset selected is obtained from Protein Data Bank (PDB). Proteins considered from two huge varieties of protein families: The Ras family and the protein kinase family are used. The PDB ID used from the Ras family is 121P, 1QRA, 1CTQ, 1P2S, 1AGP, 4DSN. Also, the PDB ID used from the protein kinase family is 4UJ1, 2PZI.

3.3. Feature Engineering

The relevant features are extracted in the pre-processing stage. All these features have been used in most of the research techniques and the same techniques as used by combining these features and converting them into formats like matrix and vector. This data is fed into the network which forms the SOM This research work primarily focuses on clustering a few varieties of super secondary structures including the prediction of β-Turn as well.

3.4. Model Creation—SOM

In this work, the two input protein families mainly used the Ras family and the protein kinase, family. After pre-processing step, the construction of the clustering Self-Organizing Map (SOM) for every protein family. The maps revealed their clusters, their size, and their relative place to everyone are

statistically significant and are fully converged. Some important parameter aspects must be considered during network construction (SOM) mainly for improving Performance and Accuracy. The parameters used in SOM are: Map size = 200, Radius = 5, Traininglength = 6000, sampling rate = 1/100 per second and Neighbourfunction = Gaussian. The SOM technique is analyzed with a neural net, popsom packages with R programming.

3.5. Average Accuracy Error Rate

Various distance measures were tested while constructing the Self Organising Map (SOM) which includes Euclidean, Manhattan, and Chebyshev. The evaluation was done on how the performances of the SOM change with different types of distance measures. The accuracy values attained with these distance measures for a different set of network parameters have been plotted as shown in Figure 14.

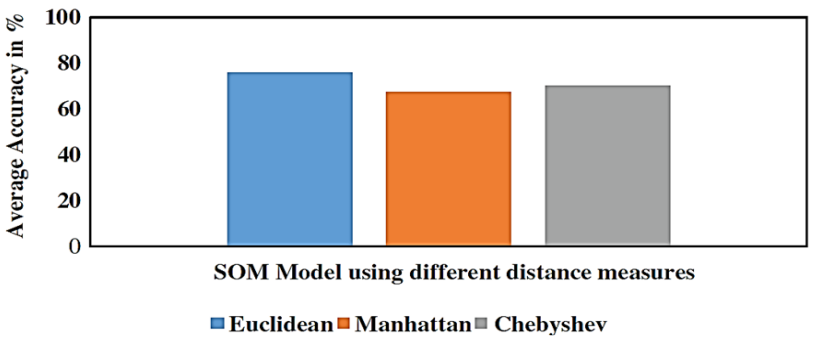


Figure 14. Accuracy Deviations concerning Different Distance Measures.

3.6. Prediction Results

The measures are calculated for identifying the cluster performance. The formulas used are given below:

Specificity is referred as Qover = [m / (m+h)]×100
Sensitivity Index is referred as Qunder = [m / (m+g)] ×100
Overall Prediction accuracy as Qtotal = m + nm +n + g + h × 100

where m and n represent the number of true +ves and true –ves, g and h represent the false +ves and false -ves. Table 4.4 shows the β-Turns prediction results obtained through Ras and protein Kinase family.

Table 4. SOM prediction results for Ras and Kinase Family

Training data	Ras family	Protein kinase family
Number of turn samples	9,567	6,398
Overall Prediction accuracy	71.4	70.6
Sensitivity	73.2	73.2
Specificity	77.3	75.5
MCC	0.56	0.55

4. Proposed Adaptive Threshold for Denoising

After obtaining a real-time image, the first step of processing it is removing the noise which is of different types like salt and pepper noise, Gaussian noise, and speckle noise using wavelet transform. The subbands of orthogonal wavelet transform are shown in figure 15. We q have proposed the spatially adaptive thresholds for denoising the images corrupted by salt and pepper (impulsive) noise and white Gaussian noise. The threshold proposed considers the local characteristics of the image under process and its properties are adapted so that the noisy coefficients can be suppressed heavily. We have considered the local threshold in (7) and modified it with a scalar coefficient and the corresponding threshold (9) works better than the local one which can be seen in the results.

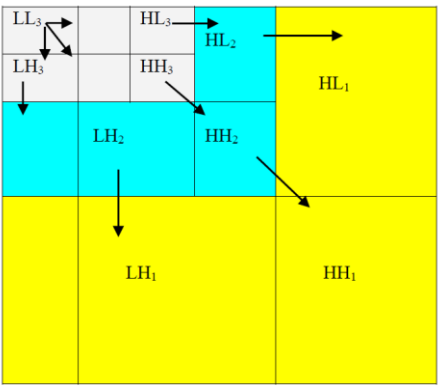


Figure 15. Subbands of the orthogonal DWT.

$$t_{adapt} = q^2\sqrt{\frac{log\, N}{N}} \tag{7}$$

we have proposed the adaptive threshold which considers the variance property of the noisy signal and with coefficient q works better compare to the existing thresholds. The proposed threshold is as

$$t_{adapt2} = q\sqrt{2 * log\, N * \sigma^2} \tag{8}$$

we have also considered the level and standard deviation of the noisy image and proposed the threshold (9)which performs better in a noisy environment.

$$t_{adapt3} = \sigma\sqrt{j * \frac{log\overleftrightarrow{\ } N}{N}} \tag{9}$$

We have also tried with following modified thresholds and obtained better results.

$$t = N\sqrt{2 * \sigma^2} \text{ and } t = q^2\sqrt{\frac{j}{N}}. \tag{10}$$

Figure 16 shows that the proposed thresholding $t = q * \sqrt{2 * log\, N * \sigma^2}$ performs better than the wavelet denoising with soft thresholding in terms of SNR, PSNR, and MSE. The proposed method removes almost all the impulsive noise as compared to the existing one and performs better both visually and quantitatively.

Figure 16. a) Noisy image with salt & pepper noise b) Wavelet denoised with soft thresholding c) Denoised with proposed threshold $t = q * \sqrt{2 * log\, N * \sigma^2}$.

SNR, PSNR, and MSE performance of different methods at the input and after filtering are tabulated in Table 1-9. Table 3.1-3.9 and Figures 3-11 show that the proposed method (i.e., proposed algorithm with spatially adaptive threshold wavelet domain filter) supersedes the wavelet denoising (WD) on the average for all the images.

Table 5. SNR, PSNR, and MSE result for the images of Figure 3

SNR(dB)			PSNR(dB)			MSE(dB)		
Noisy	Universal	Proposed	Noisy	Universal	Proposed	Noisy	Universal	Proposed
29.96	29.98	31.55	-10.7	-9.79	-8.09	10.09	10.04	7.006

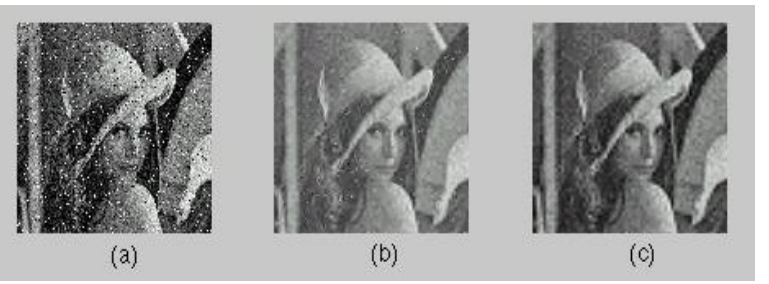


Figure 17. a) Noisy image with gaussian plus salt & pepper noise b) Wavelet denoised with hard thresholding c) Denoised with proposed threshold $t = q * \sqrt{2 * log\, N * \sigma^2}$ with large q.

Table 6. SNR, PSNR, and MSE result for the images of Figure 17

SNR(dB)			PSNR(dB)			MSE(dB)		
Noisy	Universal	Proposed	Noisy	Universal	Proposed	Noisy	Universal	Proposed
27.85	29.91	30.07	-12.8	-9.54	-9.67	16.43	10.22	9.85

Figure 18. a) Noisy image with Gaussian noise b) Wavelet denoised with hard thresholding c) Denoised with proposed threshold.

Table 7. SNR, PSNR, and MSE result for the images of Figure 18

SNR(dB)			PSNR(dB)			MSE(dB)		
Noisy	Universal	Proposed	Noisy	Universal	Proposed	Noisy	Universal	Proposed
31.40	33.67	34.66	-9.27	-5.43	-4.35	7.24	4.30	3.42

Figure 19. a) Noisy image with salt and pepper noise b) Wavelet denoised with hard thresholding c) Denoised with proposed threshold $t = q * \sqrt{2 * \sigma^2 * log\, N}$ With small q.

Table 8. SNR, PSNR, and MSE result for the images of Figure 19

SNR(dB)			PSNR(dB)			(dB)		
Noisy	Universal	Proposed	Noisy	Universal	Proposed	Noisy	Universal	Proposed
29.85	29.93	31.65	-10.8	-9.81	-7.84	10.36	10.17	6.84

Figure 20. a) Noisy image with Gaussian noise b) Wavelet denoised with local threshold $t = q \sqrt{\frac{log}{}}$i**c)** Denoised with proposed threshold $t = \sigma \sqrt{j * \frac{log\, N}{N}}$.

Table 9. SNR, PSNR, and MSE result for the images of Figure 20

SNR(dB)			PSNR(dB)			MSE(dB)		
Noisy	Local	Proposed	Noisy	Local	Proposed	Noisy	Local	Proposed
32.06	34.11	35.05	-8.60	-5.74	-4.77	6.22	3.88	3.12

Figure 21. a) Noisy image with salt & pepper noise b) Wavelet denoised with universal thresholding c) Denoised with proposed threshold $t = q^2 \sqrt{\frac{log\ N}{N}}$.

Table 10. SNR, PSNR, and MSE result for the images of Figure 21

SNR(dB)			PSNR(dB)			MSE(dB)		
Noisy	Universal	Proposed	Noisy	Universal	Proposed	Noisy	Universal	Proposed
30.96	32.64	33.29	-9.71	-7.37	-6.30	8.01	5.44	4.69

Figure 22. a) Noisy image gaussian noise b) Wavelet denoised with global threshold $t = q \sqrt{\frac{j}{N}}$ c) Denoised with proposed threshold $t = q^2 \sqrt{\frac{j}{N}}$ j level, N No of pixels.

Table 11. SNR, PSNR and MSE result for the images of Figure 22

SNR(dB)			PSNR(dB)			MSE(dB)		
Noisy	Universal	Proposed	Noisy	Universal	Proposed	Noisy	Universal	Proposed
29.02	31.74	32.13	-11.6	-7.99	-7.69	12.54	6.69	6.12

Figure 23. a) Noisy image salt & pepper noise b) Wavelet denoised with universal thresholding c) Denoised with proposed threshold $t = N\sqrt{2*\sigma^2}$, $N =$ No of pixels.

Table 12. SNR, PSNR, and MSE result for the images of Figure 23

SNR(dB)			PSNR(dB)			MSE(dB)		
Noisy	Universal	Proposed	Noisy	Universal	Proposed	Noisy	Universal	Proposed
29.97	30.04	32.006	-10.7	-9.72	-7.46	10.07	9.92	6.31

Figure 24. a) Noisy image salt & pepper noise b) Wavelet denoised with universal threshold c) Denoised with proposed threshold $t = \sqrt{2*\sigma \ *\log N}$.

Table 13. SNR, PSNR, and MSE result for the images of Figure 24

SNR (dB)			PSNR (dB)			MSE (dB)		
Noisy	Universal	Proposed	Noisy	Universal	Proposed	Noisy	Universal	Proposed
29.59	29.67	31.54	-11.0	-10.07	-8.25	10.99	10.79	7.01

Conclusion

In this paper, a new scheme for the removal of impulsive (salt & pepper) noise and Gaussian white noise is implemented, in which the wavelet coefficients are thresholded using a spatially adaptive threshold. The pulse of an algorithm is the use of a spatial adaptive threshold that considers the image statistics and performs accordingly. The proposed algorithm is tested on 2-D images to find which works better for the removal of salt and pepper as well as white Gaussian noise compared to the wavelet denoising algorithm with universal and local thresholds. 2-D image Lenna is used for testing the proposed method. The noisy test images are obtained by adding salt & pepper noise for impulse noise removal and additive Gaussian white noise to test the performance of the proposed method for the removal of Gaussian noise. The results are compared with those obtained by wavelet denoising algorithm with universal, local, and global thresholds. Daubechies wavelet basis with three vanishing moments and two finest levels are used to obtain the wavelet coefficients. After denoising, a translation and rotation invariant wavelet-based feature extractor for the efficient classification of chromosomes has been discussed. The results and observations are:

I. The application of the feature extractor results in a 98% recognition rate.
II. The minimum distance classifier requires the experimental computation of tolerance limit values for efficient recognition of the chromosomes.

References

[1] Marques de Sa, *Pattern Recognition*, Springer Verlag, 2002.
[2] Graham J and Piper J, Automatic Karyotyping Analysis, *Methods in Molecular Biology*, 29, 141-185, 1994.
[3] Detcheva E, Object-oriented approach to the design of CAIS-an an intelligent system for chromosome analysis and classification, *Proc. 21st Spring Conf. of the Union of Bulgarian Mathematicians*, 226-230, 1992.

[4] Detcheva E, Heuristic knowledge-based chromosome classification, *Proc. 1st National Conf. on Informatics*, 85-89, 1994.

[5] Jennings AM and Graham J, A Neural Network approach to automatic chromosome classification, *Phys. Med. Biol*., 38, 959-970, 1993.

[6] Lerner B, Guterman H, Dinstein I and Y Romem, Human chromosome classification using Multi-layer Perceptron Network, *Intl. J. Neural Systems*, 6(3), 359-370, 1995.

[7] Lerner B, Towards a completely automatic Neural Network-based human chromosome analysis, *IEEE Trans. On Systems, Man and Cybernetics*, 28(4), 544-552, 1998.

[8] Mousavi P, Fels S, Ward R, and Lansdorp P, Classification of homologous human chromosomes using Mutual Information Maximization, *Proc. Intl. Conf. on Image Processing*, 2, 845-848, 2001.

[9] Cho Jong Man, Chromosome classification using Backpropagation Neural Networks, *Engineering in Medicine and Biology Magazine*, IEEE, 28-3, 19(1), 2000.

[10] Wu Q and Castleman KR, Automated chromosome classification using wavelet-based pattern descriptors, *Proc. 13th IEEE Symposium on Computer-based medical systems*, 189-194, 2000.

[11] Tang YY, Cheng HD and Suen CY, Transformation ring projection algorithm and its VLSI implementation, *Intl. J. Pattern Recognition and Artificial Intelligence*, 5(1,2) 25-56, 1991.

[12] Tang YY, Yang LH, Liu J, and Ma H, *Wavelet theory and its application to Pattern Recognition*, World Scientific, 2000.

[13] Mukundan R and Ramakrishnan KS, *Moment functions in Image Analysis*, World Scientific Publication, 1998.

[14] Rajasekaran S and Vijayalakshmi Pai GA, Simplified Fuzzy ARTMAP as a Pattern Recognizer, *ASCE Journal of Computing in Civil Engineering*, Vol. 14, No. 2, pp. 92-99, April 2000.

[15] Rajasekaran S and Vijayalakshmi Pai GA, Image recognition using Simplified Fuzzy ARTMAP augmented with a moment-based feature extractor, *Intl. Journal of Pattern Recognition and Artificial Intelligence*, Vol. 14, No. 8, 1081-1095, 2000.

[16] Rajasekaran S, Vijayalakshmi Pai GA and Subbarayudu P, Rock identification using Simplified Fuzzy ARTMAP augmented with a moment-based feature extractor, *Technology*, pp. 45-50, 1999.

[17] Jayakumar S, *Chromosome classification using Kohonen's Self-organizing feature map*, Master of Computer Application thesis, Department of Computer Applications, PSG College of Technology, 2002.

[18] Sivagamasundari, G, and Dr. Latha Parthiban, "Clustering Protein Super Secondary Structure with Artificial Neural Network Based Self Organizing Map Machine Learning Approach Using In-Memory Computing Environment" *Journal of Computational and Theoretical Nano-science* Vol. 17, 1–3, 2020.

[19] Shine Babu, D, Dr. Latha Parthiban, G. Sivagamasundari, "Motif Structure Prediction in Distributed Framework using Machine Learning Algorithms" *International Journal of Recent Technology and Engineering (IJRTE)* ISSN: 2277-3878, Volume-7, Issue-5S4, February 2019.

[20] Smita C Thomas, Latha Parthiban (2018) Investigation of drug-induced adverse effects and Disproportionality Analysis of Acetaminophen and Ibuprofen, *International Journal of Pure and Applied Mathematics*, Volume 118 No. 7 2018, 667-674.

[21] Sriramakrishnan, G, Latha Parthiban (2016) Data mining techniques for finding serious Adverse Events, *Journal of Chemical and Pharmaceutical Sciences*, Vol 9, No 4, pp.1-5.

Chapter 7

Quantum Based Dynamic Clustering of Pharmacovigilance Data

S. Nithinsha[1,*] and D. Vinod[2]

[1]Department of CSE, Saveetha School of Engineering,
SIMATS, Chennai, Tamil Nadu, India
[2]Department of Cloud and Edge Analytics,
Institute of Computer Science and Engineering,
Saveetha School of Engineering, Chennai,
Tamil Nadu, India

Abstract

Data clustering can be achieved in both static methods and dynamically in run time. In the static method, clustering consumes more time when new documents are added to the already existing clusters. To reduce the time and improve the efficiency of clustering, a dynamic document algorithm has been proposed for clustering pharmacovigilance data. In this work, standard clustering methods like k-means and their variants are compared with the proposed technique, and the experimental results with the proposed algorithm are found to work better.

Keywords: dynamic clustering, pharmacovigilance data, k-means, Maximum Resemblance Data Labeling

[*] Corresponding Author's Email: ayshaas2017@gmail.com.

In: Frontiers in Quantum Computing
Editor: R. Anandan
ISBN: 978-1-68507-816-4
© 2022 Nova Science Publishers, Inc.

1. Introduction

There are many clustering algorithms available in the literature and some common techniques like k-means are discussed. The k-means objective function is frequently used in document clustering, in part due to the popularity and (perceived) speed of Lloyd's.

1.1. K-Means Partitioning

Several other algorithms have been proposed that also attempt to minimize this objective function, including bisecting k-means and simple k-means [1, 2]. The spectral technique also has a k-means objective reconstitution technique.

Another reason for the popularity of k-means is a property of vector representation [3]. If the vector dimensions represent topics and the values are binary, then the square of the distance between any two is the number of topics on which it rests [4].

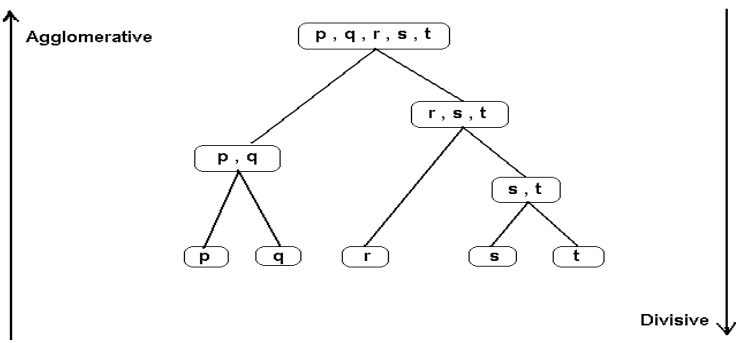


Figure 1. Hierarchical clustering.

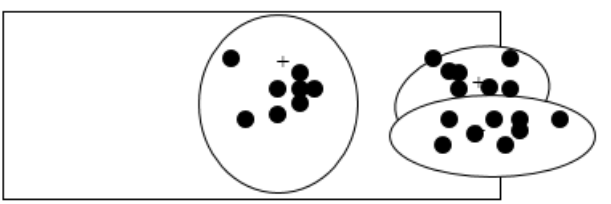


Figure 2. K-means partitioning technique.

Furthermore, their dot product is the number of topics on which they agree and explains the intuition behind using Lloyd's algorithm with distances based on cosine similarity [5, 6].

Figure 2 shows the bisecting k-means algorithm where the single cluster is bisected into a set of k-clusters.

2. Dynamic Document Clustering

Figure 3 shows the overall schematic of the proposed work showing the various steps in the process.

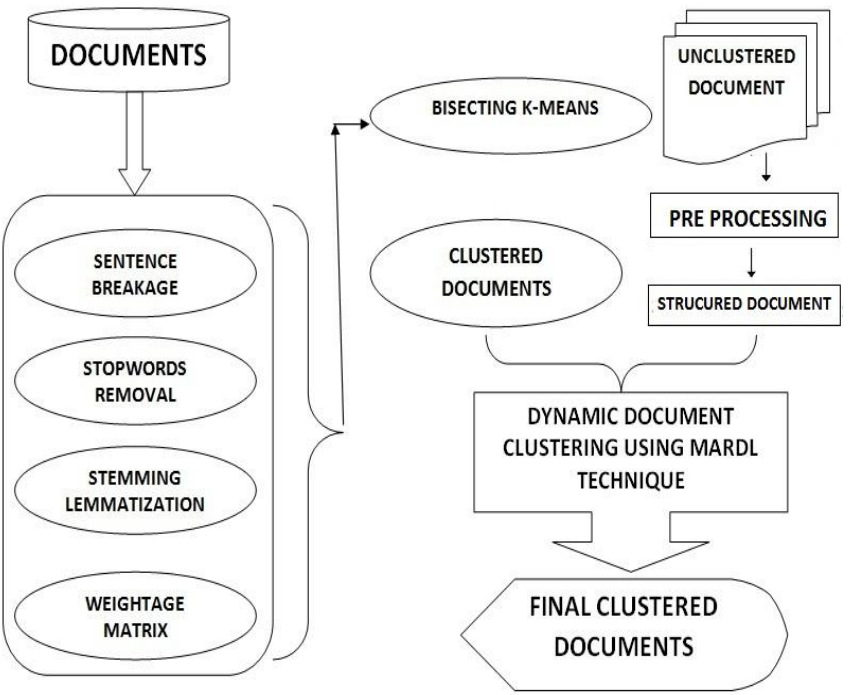


Figure 3. Overall process of proposed work.

2.1. Description of Proposed Dynamic Clustering Algorithm

Figure 4 provides a clear step-by-step process of the algorithm proposed.

3. Proposed Clustering Using Maximum Resemblance Data Labeling (MARDL) Technique

The steps in the algorithm are as follows:

1. Using bisecting K-means algorithm, cluster the documents and form K-clusters.
   $C=\{C_1,C_2,C_3,\ldots..C_k)$ $1<=i<=k$ $C_i=i^{th}$ cluster

2. Get a Sample from each cluster in C and each sample consists of a set of documents.

   $S=\{S_1,S_2,S_3,\ldots.S_k\}$ where S_k is a sample of cluster C_i.

   $d_j=\{d_1,d_2,d_3,\ldots\ldots d_j\}$ where $d_j=1/3$ of the no.of documents of sample S_j

3. If new documents arrive in the system to cluster, preprocess the documents as a set of words for each sentence in the document.

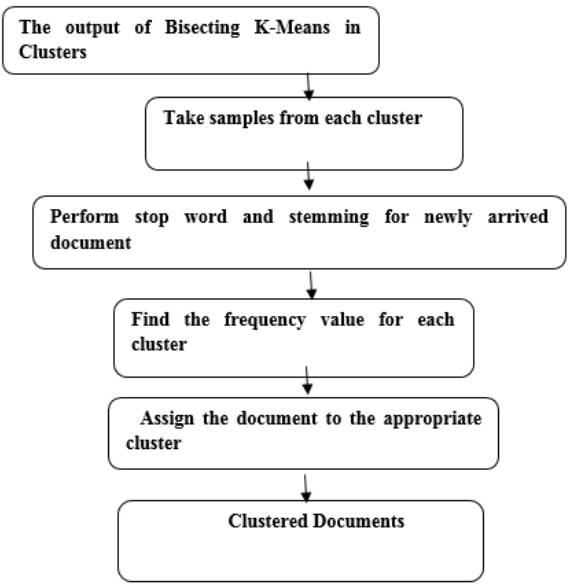


Figure 4. Steps in the proposed clustering algorithm.

doc=(doc_1,doc_2,doc_3….doc_j) doc_j=a newly arrived document

4. For every document d_1 there is a list of sentences d1={T_1,T_2,…..T_n}.calculate the score matrix.

	doc_1	doc_2	doc_3 ….	doc_j
T_1	T_1doc_1	T_1doc_2	T_1doc_3	T_1doc_j
T_2	T_2doc_1	T_2doc_2	T_2doc_3	T_2doc_j
T_3	T_3doc_1	T_3doc_2	T_3doc_3	T_3doc_j
.				
.				
Tn	T_ndoc_1	T_ndoc_2	T_ndoc_3	T_ndoc_j

where T_ndoc_j – influence of sentence T_n in document doc_j.

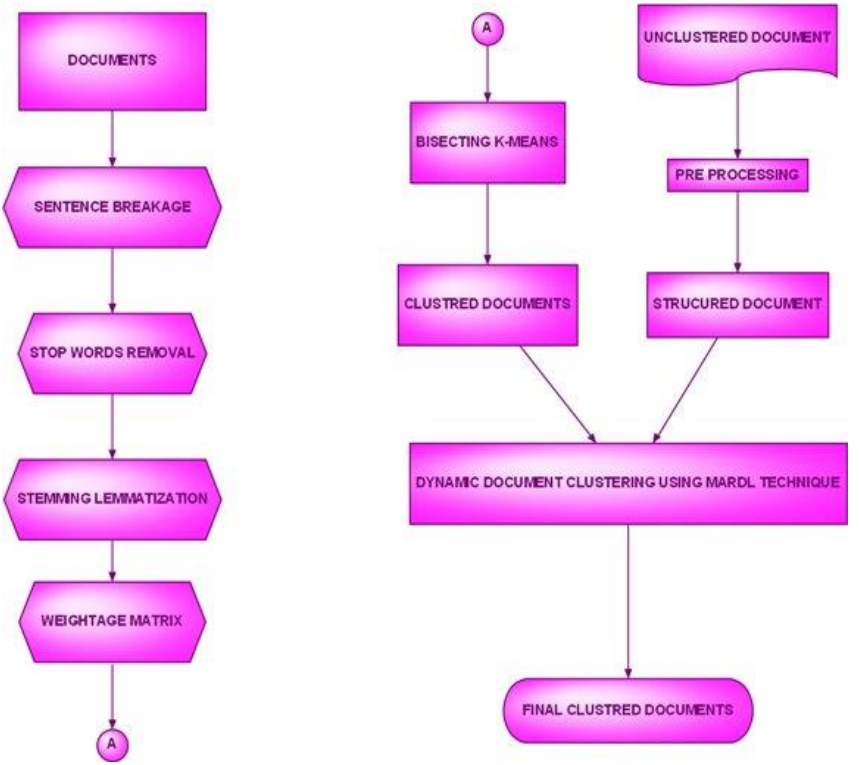


Figure 5. Data flow diagram of proposed work.

5. Calculate the weight-age of each sentence in document doc_j using column-wise summation.

$$\text{Sentence Importance Calculation } (K_i) = \sum_{i=1}^{n} T_i \, doc_j$$

$$\text{Also calculate the weight-age of } doc_j \text{ in sample } S_{i,I} = \sum_{k=1}^{j} \left\{ \sum_{i=1}^{n} T_i \, doc_j \right\}$$

$$I$$

   Frequency value of doc_j in sample S_i, F = Number of documents in sample S_i
6. Repeat step 5 for all samples and calculate the Frequency value of the new document doc_j, with each sample. From the series of frequency values obtained, the maximum is termed as F_{max}.
7. Include the doc_j (newly arrived document) to i^{th} cluster
8. If the calculated F_{max} is less than the threshold value (T_{max}), then the newly arrived document forms a new cluster. After a series of experiments, T_{max} is arrived at.

The data flow diagram of the proposed work is shown in figure 5. The existing work bisecting K-Means [7, 8] and proposed dynamic clustering algorithm have been implemented using Java Eclipse software.

3.1. Implementation Code

3.1.1. K-Means Clustering

The Java implementation of k-means algorithm is given below

```
import java.io.*;
import java.util.*;
import java.lang.*;
class cluster
{
int c,n,i,j,kn = 0;
int[] obj,doc;
```

```
int[][] cl;
static int docs[] = new int[500];// number of sample docs taken for each
cluster
public HashMap<Long,ArrayList<String>>  docSampleIdentify()
{
int clustno=1;
String str,s;
int n;
StringTokenizer st;
FileInputStream fstream;
DataInputStream in;
FileOutputStream stout,stout1;
PrintStream stp,stp1;
HashMap<Long,ArrayList<String>>sampledDataHashMap=new
HashMap<Long,ArrayList<String>>();
try
{
String file2="output/newdocAssignment.txt";
fstream = new FileInputStream(file2);
in = new DataInputStream(fstream);
while (in.available() !=0)
{
str=in.readLine();
str=in.readLine();
st = new StringTokenizer(str);
s=st.nextToken();
j++;
n=Integer.parseInt(s);
System.out.println("cluster:" + (j));
System.out.println("Number of documents:"+n);
if(n>1)
{
docs[j]=n/3;
}
else
{
docs[j]=n;
}
```

```
System.out.println("Number of samples should taken:"+ docs[j]);
str=in.readLine();
StringUtil ob = new StringUtil();
ArrayList<String> docIds=  ob.stringToArrayList(str);
ArrayList<String> mergeSamplelist = new ArrayList<String>();
for ( int i=0;i<docs[j];i++){
// code to sample the ramdom files
Random r = new Random();
String sampleDocId = docIds.get(r.nextInt(docIds.size()));
System.out.println("sampleDocId:");
Newstopword newstopword=new Newstopword();
mergeSamplelist.addAll(wordStringList);
}
System.out.println( );
sampledDataHashMap.put(new Long(j), mergeSamplelist);
file2="output/sampledDataHashMap.txt";
stout = new FileOutputStream(file2);
stp = new PrintStream(stout);
int clusterCount =0;
for(Long sampledKey:sampledDataHashMap.keySet() ){
clusterCount++;
stp.println("CLUSTER"+clusterCount);
ArrayList<String>mergeList=(ArrayList<String>)sampledDataHashMa
p.get(sampledKey)
for (String strings:mergeList){
stp.println(strings);
}
}
}
FileOutputStream("output/sampledDataHashMap.txt");
}
catch(Exception e)
{
//System.out.println("\n\nError in words function");
}
return sampledDataHashMap;
}
}
```

3.1.2. Bisecting K-Means

The java implementation of bisecting k-means algorithm is given below

```
public class prop
{
cluster ob = new cluster();
FileInputStream fstream;
DataInputStream in;
Scanner inp;
FileOutputStream stout;
PrintStream stp;
int n;
int a[]=new int[10];
//step 01 sampling with output of existing Clusters data
public  HashMap<Long,ArrayList<String>>  sampleCreation()
{
HashMap<Long,ArrayList<String>> sampledDataHashMap =null;
try
{
sampledDataHashMap= ob.docSampleIdentify();
}
catch(Exception e)
{
System.out.println("\n\nError in start function");
}
return sampledDataHashMap;
}
// step 02 new file stopword and stemming
public void stopword(String s)
{
try
{
Newstopword  ob = new Newstopword();
ob.removal(s);
}
catch(Exception e)
{
System.out.println("\n\nError in stopword function");
}
}
```

```
public ArrayList<String> stemming(String a)
{
ArrayList<String> list= new ArrayList<String>();
try
{
Stemmer ob;
String str,s1,s,file1,file2,file3;
StringTokenizer st;
String n;
n=a;
file1="dynstop/"+ n +".txt";
fstream = new FileInputStream(file1);
in = new DataInputStream(fstream);
file2="dynstem/"+n+".txt";
file2="stem/"+n+".txt";
stout = new FileOutputStream(file2);
stp = new PrintStream(stout);
while (in.available() !=0)
{
ob = new Stemmer();
s=ob.stem(s1);
stp.println(s);
list.add(s);
}
}
catch(Exception e)
{
System.out.println("\n\nError in stemming function");
}
return list;
}
public static void main(String args[])
{
cluster ob = new cluster();
try
{ prop obj = new prop();
String docAddedtoCateg=null;
int j=1;
for(Long sampledKey:sampledDataHashMap.keySet() ){
```

```
Long occurance=0L;
for(String wordFromNewdoc:newDocWordList){
for (String strings:mergeList){
if (strings.compareTo(wordFromNewdoc)== 0){
occurance ++;
}
}
}
System.out.println("cluster :"+sampledKey);
System.out.println("Total Occurance :"+ occurance);
System.out.println("Number of samples taken :"+ ob.docs[j]);
occurance=occurance/ob.docs[j];
j++;
docWtPerCluster.put(sampledKey, occurance);
System.out.println("Average occurance :" + occurance);
System.out.println("    ") ;
}
//System.out.println("documents"+j);
Long highestOccrClusterId=null;
Long temp=0L;
for (Long keys:docWtPerCluster.keySet()){
if(docWtPerCluster.get(keys)>temp){
temp=docWtPerCluster.get(keys);
highestOccrClusterId=keys;
}
}
if(temp>20)
System.out.println("highestOccrClusterId :"+highestOccrClusterId);
}
else
{
System.out.println("it shud form to a new cluster");
}
String file2 ,str,s;
StringTokenizer st;
FileInputStream fstream;
DataInputStream in;
FileOutputStream stout;
PrintStream stp;
```

```
double d,td,tw,nd,nw;
file2="output/newdocAssignment.txt";
// reading the corresponding cluster line
fstream = new FileInputStream(file2);
in = new DataInputStream(fstream);
StringBuilder builder=new StringBuilder();
s="";
if(temp>20)
{
while (in.available() !=0)
{
builder.append(str);
builder.append('\n');
if(str.compareTo("CLUSTER"+highestOccrClusterId+":") ==0) {
str=in.readLine();// returns the number of doc in the cluster
//now we will increment the value
int count =Integer.parseInt(str);
count++;
System.out.println(count);
builder.append(count);
builder.append('\n');
str=in.readLine();
docAddedtoCateg=str+" "+newDocId;
builder.append(docAddedtoCateg);
builder.append('\n');
System.out.println(count+"  "+docAddedtoCateg);
}
}
stout = new FileOutputStream(file2);
stp = new PrintStream(stout);
stp.println(builder.toString());
}
else
{
builder.append(str);
builder.append('\n');
}
docAddedtoCateg="CLUSTER"+j+":";
builder.append(docAddedtoCateg);
```

```
builder.append('\n');
int count1=1;
builder.append(count1);
builder.append('\n');
builder.append(newDocId);
stout = new FileOutputStream(file2);
stp = new PrintStream(stout);
stp.println(builder.toString());
j=j+1;
}
}
catch(Exception e)
{
System.out.println("\n\nError in main function");}}}
```

3.1.3. Weight Matrix

The code for calculating the weight matrix is given below

```
public class WeightMatrix
{
FileInputStream fstream;
DataInputStream in;
FileOutputStream stout;
PrintStream stp;
int n;
String[] nDupe;
/*
public static void main(String[] args)
{
WeightMatrix ob = new WeightMatrix();
ob.start();
}
*/
public void start(int x)
{
try
{
String str,s,s1,s2,file1,file2;
StringTokenizer st;
int i,f;
```

```
n=x;
//n=3;
s="";
for(i=1;i<=n;i++)
{
file1="weight/"+String.valueOf(i)+".txt";
fstream = new FileInputStream(file1);
while (in.available() !=0)
{
s1=st.nextToken();
s2=st.nextToken();
s=s+" "+s1;
}
}
String[] wDupes=s.split(" ");
String noDupes = "";
for (f = 0; f < wDupes.length; f++)
{
if (noDupes.indexOf(wDupes[f]) < 0)
{
// -1 otherwise, if not in noDupes
noDupes = noDupes+" "+wDupes[f];
}
}
//String[] nDupe=noDupes.split(" ");
nDupe=noDupes.split(" ");
file2="outdocs/words.txt";
stout = new FileOutputStream(file2);
stp = new PrintStream(stout);
for(f=1;f<nDupe.length;f++)
{
stp.println(nDupe[f]);
}
calc();
}
catch(Exception e)
{
System.out.println("\n\nError in start function");
```

```
}
}
public void calc()
{
try
{
String str,s1,s2,s3,s4,file1,file2;
StringTokenizer st;
int i,j,k,wn,a;
double x;
String[] wd,wt;
for(i=1;i<=n;i++)
{
file1="weight/"+String.valueOf(i)+".txt";
fstream = new FileInputStream(file1);
in = new DataInputStream(fstream);
s3="";
s4="";
while (in.available()!=0)
{
s1=st.nextToken();
s3=s3+" "+s1;
s2=st.nextToken();
s4=s4+" "+s2;
}
wd=s3.split(" ");
wt=s4.split(" ");
file2="outdocs/"+String.valueOf(i)+".txt";
stout = new FileOutputStream(file2);
stp = new PrintStream(stout);
wn=nDupe.length;
for(k=1;k<wn;k++)
{
a=0;
for(j=0;j<wd.length;j++)
{
if(nDupe[k].compareTo(wd[j])==0)
{
break;
```

```
}
a++;
}
if(a<wd.length)
{
stp.println(nDupe[k]+" "+wt[j]);
}
else
{
stp.println(nDupe[k]+" 0.0");
}
}
}
}
catch(Exception e)
{
System.out.println("\n\nError in calc function");
}
}
}
```

3.1.4. Proposed Algorithm

The Java implementation of the proposed algorithm is given below

```
public class prop
{
cluster ob = new cluster();
FileInputStream fstream;
DataInputStream in;
Scanner inp;
FileOutputStream stout;
PrintStream stp;
int n;
int a[]=new int[10];
//step 01 sampling with output of existing Clusters data
public  HashMap<Long,ArrayList<String>>  sampleCreation()
{
HashMap<Long,ArrayList<String>> sampledDataHashMap =null;
try
```

```
{
sampledDataHashMap= ob.docSampleIdentify();
}
catch(Exception e)
{
System.out.println("\n\nError in start function");
}
return sampledDataHashMap;
}
// step 02 new file stopword and stemming
public void stopword(String s)
{
try
{
Newstopword  ob = new Newstopword();
ob.removal(s);
}
catch(Exception e)
{
System.out.println("\n\nError in stopword function");
}
}
public ArrayList<String> stemming(String a)
{
ArrayList<String> list= new ArrayList<String>();
try
{
Stemmer ob;
String str,s1,s,file1,file2,file3;
StringTokenizer st;
String n;
n=a;
file1="dynstop/"+ n +".txt";
fstream = new FileInputStream(file1);
in = new DataInputStream(fstream);
file2="dynstem/"+n+".txt";
file2="stem/"+n+".txt";
stout = new FileOutputStream(file2);
stp = new PrintStream(stout);
```

```
while (in.available() !=0)
{
s1=st.nextToken();
ob = new Stemmer();
s=ob.stem(s1);
stp.println(s);
list.add(s);
}
}
catch(Exception e)
{
System.out.println("\n\nError in stemming function");
}
return list;
}
public static void main(String args[])
{
cluster ob = new cluster();
try
{
prop obj = new prop();
sampledDataHashMap =  obj.sampleCreation();//step1
obj.stopword(args[0]);//step 2
ArrayList<String> newDocWordList = obj.stemming(args[0]);//step 2
//step 03 search the doc contents in the sampling
HashMap<Long, Long> docWtPerCluster=new HashMap<Long, Long>
();
// new doocument id
int newDocId=Integer.parseInt(args[0]);//
String docAddedtoCateg=null;
int j=1;
for(Long sampledKey:sampledDataHashMap.keySet() ){
Long occurance=0L;
for(String wordFromNewdoc:newDocWordList){
ArrayList<String>mergeList=(ArrayList<String>)
for (String strings:mergeList){
if (strings.compareTo(wordFromNewdoc)== 0)
{
occurance ++;
```

```
}
}
}
System.out.println("cluster :"+sampledKey);
System.out.println("Total Occurance :"+ occurance);
System.out.println("Number of samples taken :"+ ob.docs[j]);
occurance=occurance/ob.docs[j];
j++;
docWtPerCluster.put(sampledKey, occurance);
System.out.println("Average occurance :" + occurance);
System.out.println("    ") ;
}
//System.out.println("documents"+j);
Long highestOccrClusterId=null;
Long temp=0L;
for (Long keys:docWtPerCluster.keySet()){
if(docWtPerCluster.get(keys)>temp){
temp=docWtPerCluster.get(keys);
highestOccrClusterId=keys;
}
}
if(temp>20)
{
System.out.println("highestOccrClusterId :"+highestOccrClusterId);
}
else
{
System.out.println("it shud form to a new cluster");
}
//step 04 assign the new file to an existing cluster or create one.
String file2 ,str,s;
StringTokenizer st;
FileInputStream fstream;
DataInputStream in;
FileOutputStream stout;
PrintStream stp;
double d,td,tw,nd,nw;
file2="output/newdocAssignment.txt";
// reading the corresponding cluster line
```

```
fstream = new FileInputStream(file2);
in = new DataInputStream(fstream);
StringBuilder builder=new StringBuilder();
s="";
if(temp>20)
{
while (in.available() !=0)
{
builder.append(str);
builder.append('\n');
if(str.compareTo("CLUSTER"+highestOccrClusterId+":") ==0) {
str =in.readLine();// returns the number of doc in the cluster
//now we will increment the value
int count =Integer.parseInt(str);
count++;
System.out.println(count);
builder.append(count);
builder.append('\n');
str =in.readLine();
docAddedtoCateg=str+" "+newDocId;
builder.append(docAddedtoCateg);
builder.append('\n');
System.out.println(count+"  "+docAddedtoCateg)}}
stout = new FileOutputStream(file2);
stp = new PrintStream(stout);
stp.println(builder.toString());
//obj.clustering();
}
else
{
while (in.available() !=0)
{
str=in.readLine();
builder.append(str);
builder.append('\n');
}
docAddedtoCateg="CLUSTER"+j+":";
builder.append(docAddedtoCateg);
builder.append('\n');
```

```
int count1=1;
builder.append(count1);
builder.append('\n');
builder.append(newDocId);
stout = new FileOutputStream(file2);
stp = new PrintStream(stout);
stp.println(builder.toString());
j=j+1;
}
}
catch(Exception e)
{
System.out.println("\n\nError in main function");
}
}
}
```

4. Results and Discussion

Figure 6 shows the pre-processing stage of every document using k-means.

4.1. Bisecting K-Means

Figure 6 shows the documents being grouped into 5 clusters. Each document belongs to one or the other category of domains including computer, medicine, mathematics, thermodynamics, etc. [7-9].

These clusters further act as data points for the next level of hierarchy which in turn iterates until one global cluster is achieved.

It is observed that the quality of the cluster increases when the number of clusters increases.

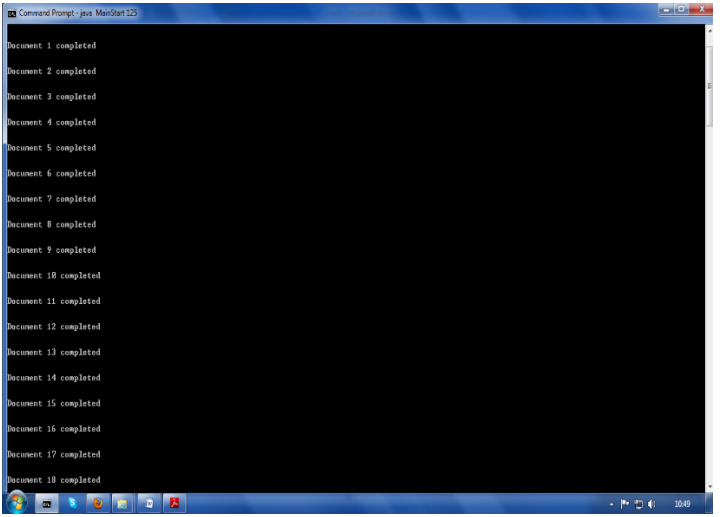


Figure 6. Pre-processing stage.

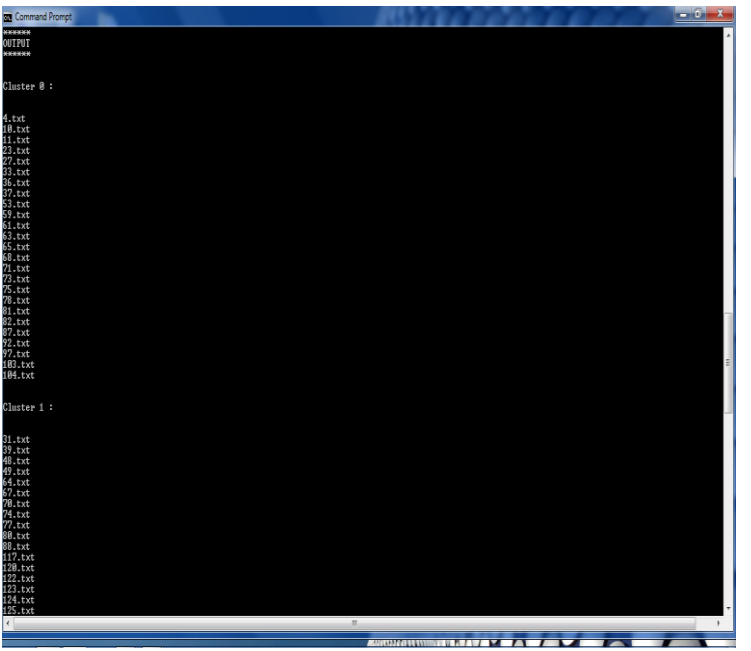


Figure 7. Bisecting K-means algorithm cluster formation.

Figure 7 provides snapshots of Bisecting K-means cluster formation. Figure 8 shows the bisecting K-means algorithm implementation.

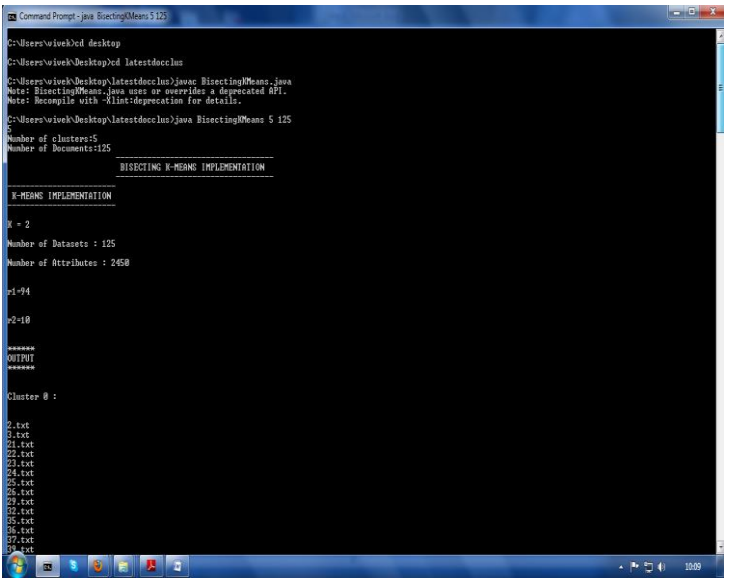


Figure 8. Bisecting K-means algorithm implementation.

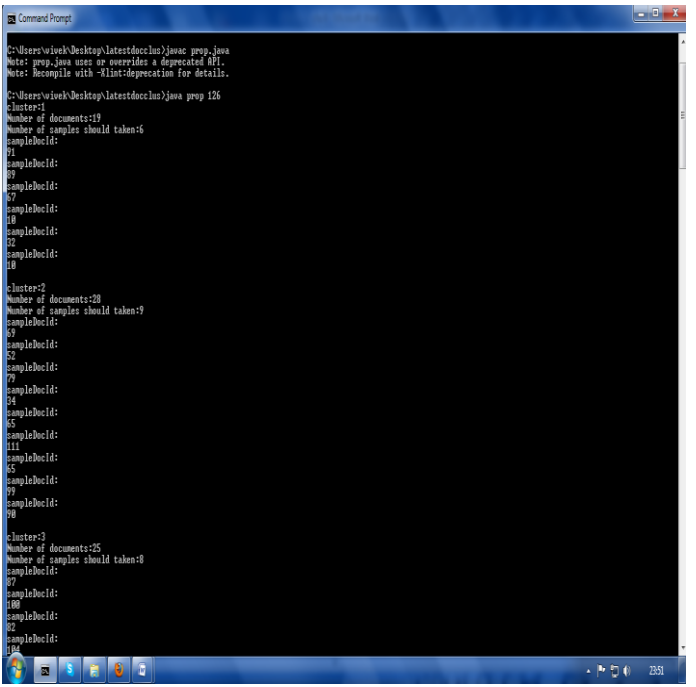


Figure 9. Pharmacovigilance document cluster formation.

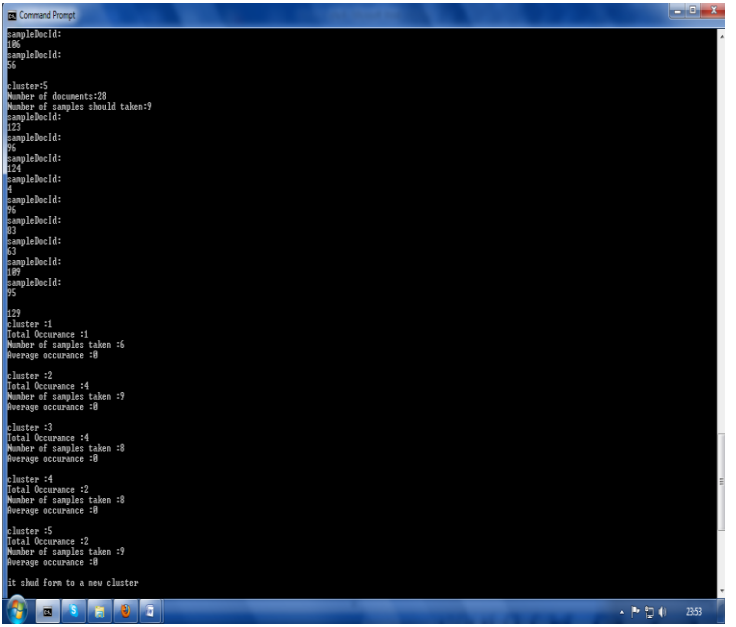


Figure 10. Document clustering using the proposed method.

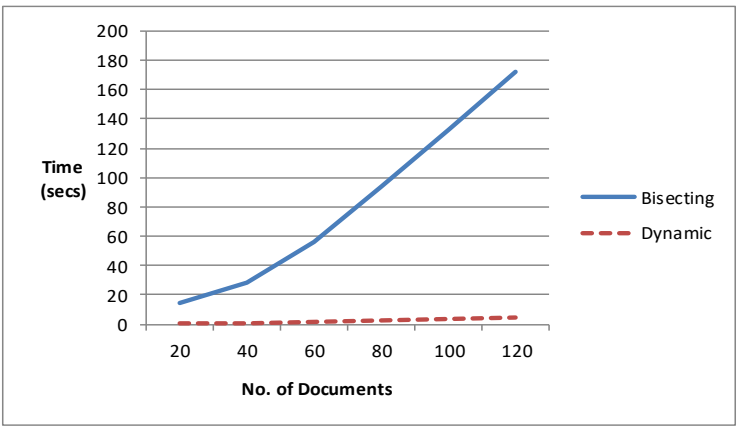


Figure 11. Performance of Static versus Dynamic Algorithm in Document Clustering Algorithm concerning Time Analysis.

4.2. Proposed Dynamic Algorithm

The first step in the proposed algorithm is the formation of a cluster.

4.2.1. Dynamic Algorithm for Forming New Cluster

In the dynamic document clustering algorithm, the frequency weight-age of every document is calculated.

Figure 9 shows the Pharmacovigilance document cluster formation and figure 10 shows the clustering output result of the proposed method. Thus the newly arrived document doc_j appears not to be similar to any of the existing clusters and so the average occurrence turns to be zero in all cases. So it forms a separate cluster by popping out a message "it should form a new cluster" as shown in the above figure 10.

4.3. Performance Analysis

4.3.1. Comparison Concerning the Time of Static Bisecting K-Means Algorithm and Proposed Dynamic Document Concerning the Time

The performance of the clustering algorithm is measured based on the clustering speed, purity, inter-cluster similarity, and intra-cluster similarity. Experiments were done to analyze the performances of Static Bisecting K-means algorithm and Proposed Dynamic Document Clustering Algorithm Figure 11 shows that dynamic clustering algorithm has better performance.

4.3.2. Static Bisecting K-Means Algorithm and Proposed Dynamic Algorithm in Purity

Purity is a metric analyzing the performance of algorithms and figure 12 show the performance of algorithms.

$$\textbf{Purity} = \frac{\text{Max(documents belong to same class in a cluster)}}{\text{Total no of documents in that cluster}}$$

Bisecting k-means

No of documents: 60			
No of clusters: 3			
Original cluster:	15	20	25
Cluster formed:	14	19	27
Max no of docs in Each cluster	13	14	22

Purity= (13+14+22)/ (14+19+27) =49/60 = **0.82**

Dynamic Document Clustering Algorithm

No of documents: 60
No of clusters: 3

Original cluster:	15	20	25
Cluster formed:	14	18	28
Max no of docs in Each cluster	13	14	23

Purity= (13+14+23)/ (14+18+28) =50/60 = **0.85**

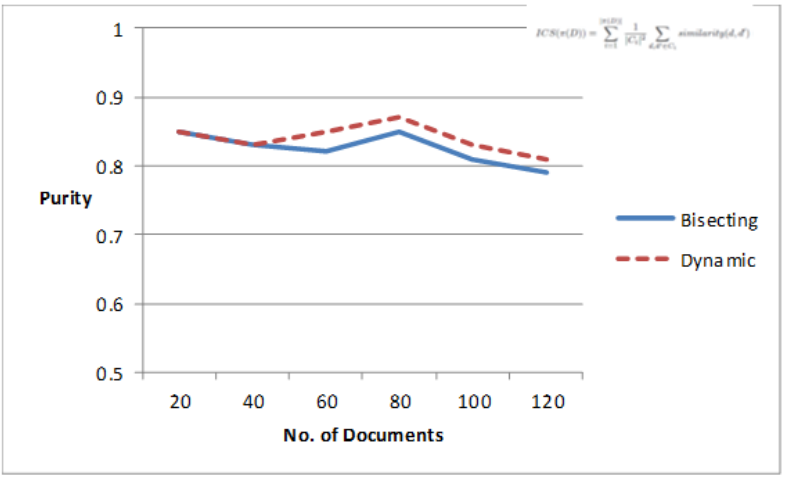


Figure 12. Performance of Static versus dynamic Algorithm in Purity.

4.3.3. Static Bisecting K-Means Algorithm and Proposed Dynamic Algorithm in Intracluster Similarity

Intra-cluster similarity measures the nearness of data objects present in a cluster.

Intra cluster similarity
Ci- no of documents
d,d'- documents within the cluster

Bisecting k-means
No of documents: 60
No of clusters: 3

Original cluster:	15	20	25
Cluster formed:	14	19	27

The first cluster formed as 13 in the first cluster and 2 in the second cluster.The second cluster formed as 1 in the first cluster 14 in the second cluster and 5 in the third cluster.

The third cluster formed as 3 in the second cluster and 22 in the third cluster.

Intra-cluster similarity=

$\{(13^2+1^2/14^2)+(2^2+14^2+3^2/19^2)+(5^2+22^2/27^2)\}/3=(0.96+0.57+0.69)/3=$ 2.26/3 = **0.742**

Dynamic Document Clustering Algorithm

No of documents: 60

No of clusters: 3

Original cluster:	15	20	25
Cluster formed:	14	18	28

The first cluster formed as 13 in the first cluster and 2 in the second cluster.

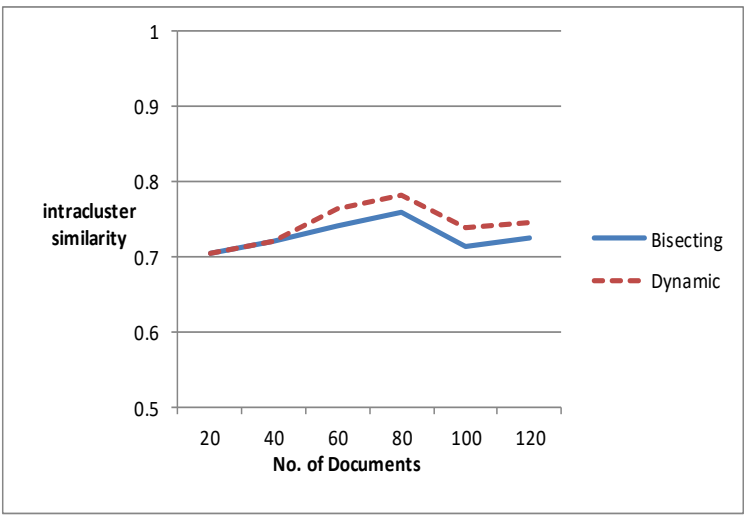


Figure 13. Performance of Static versus dynamic Algorithm in Intracluster similarity.

The second cluster formed as 1 in the first cluster 14 in the second cluster and 5 in the third cluster.

The third cluster formed as 2 in the second cluster and 23 in the third cluster.

Intra-cluster similarity=

$\{(13^2+1^2/14^2)+(2^2+14^2+2^2/18^2)+ (5^2+23^2/28^2)\}/3=$ (0.96+0.629+0.70)/3 = 2.89/3 = **0.763**

4.3.4 Static Bisecting K-Means Algorithm and Proposed Dynamic Algorithm in Inter-Cluster Similarity

Inter-cluster similarity measures how far the data objects are present between two clusters.

Inter cluster similarity

Ci,Cj- two different clusters

d,d'- documents in two different clusters

$$\sum_{i=1}^{|\pi(D)|}\sum_{j=i+1}^{|\pi(D)|}\frac{1}{|C_i||C_j|}\sum_{d\in C_i,d'\in C_j} similarity(d,d')$$

$$ECS(\pi(D)) =$$

Bisecting k-means

No of documents: 60

No of clusters: 3

Original cluster:	15	20	25
Cluster formed:	14	19	27

The first cluster formed as 13 in the first cluster and 2 in the second cluster. The second cluster formed as 1 in the first cluster 14 in the second cluster and 5 in the third cluster.

The third cluster formed as 3 in the second cluster and 22 in the third cluster.

Inter-cluster similarity =

{(13*2+1*14/14*19)+(1*5+0*22/14*27)+(14*5+3*22/27*19)}/3= (0.11+0.007+0.042)/3 = 0.159/3 = 0.053

Dynamic Document Clustering Algorithm

No of documents: 60

No of clusters: 3

Original cluster:	15	20	25
Cluster formed:	14	18	28

The first cluster formed as 13 in first cluster and 2 in second cluster. The second cluster formed as 1 in first cluster 14 in second cluster and 5 in third cluster.

The third cluster formed as 2 in second cluster and 23 in third cluster.

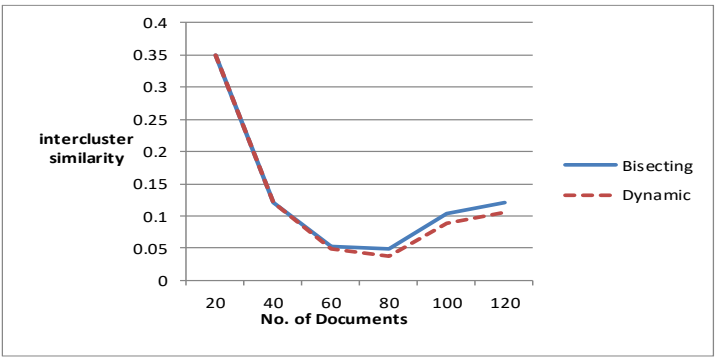


Figure 14. Performance of Static versus dynamic Algorithm in Inter cluster similarity.

Inter-cluster similarity =

{(13*2+1*14/14*18)+(1*5+0*22/14*28)+(14*5+2*23/28*18)}/3= (0.12+0.007+0.022)/3 = 0.149/3 = **0.0493**

Conclusion

The experimental results provide the overall performance of a dynamic document clustering algorithm compared to static bisecting K-means. The dynamic algorithm shows better performance than other existing algorithms in terms of speed, purity, inter and intra-cluster similarity.

References

[1] Smita C Thomas, Parthiban Latha. (2018). Investigation of drug-induced adverse effects and Disproportionality Analysis of Acetaminophen and Ibuprofen, *International Journal of Pure and Applied Mathematics*, Volume 118, No. 7, 2018, 667-674.

[2] Sriramakrishnan G, Parthiban Latha. (2016). Data mining techniques for finding serious Adverse Events, *Journal of Chemical and Pharmaceutical Sciences,* Vol 9, No 4, pp.1-5.

[3] Shaurya Jauhari and Rizvi SAM. (2014). 'Mining Gene Expression Data focusing cancer Therapeutics: A Digest' *IEEE transactions on computational biology and bioinformatics,* Volume.11, No.3.

[4] Sivagamasundari G and Dr. Parthiban Latha. (2020). "Clustering Protein Super Secondary Structure with Artificial Neural Network Based Self Organizing Map Machine Learning Approach Using In-Memory Computing Environment*". Journal of Computational and Theoretical Nano-science,* Vol. 17, 1-3.

[5] Shine Babu D, Dr. Parthiban Latha, Sivagamasundari G. (2019). "Motif Structure Prediction in Distributed Framework using Machine Learning Algorithms". *International Journal of Recent Technology and Engineering (IJRTE),* ISSN: 2277-3878, Volume-7, Issue-5S4.

[6] Kamatchi Sankar, Parthiban Latha. (2020). "Effective Pharmacovigilance with Machine Learning Algorithms Using Openvigil 2 *Solid State Technology*, Vol. 63, No. 6 (2020).

[7] Kamatchi Sankaravadivel S, Parthiban Latha. "An improved frequent item-sets mining in pharmacovigilance". *International Journal of Engineering and Advanced Technology(IJEAT),* ISSN:2277-3878, Volume-8, Issue-2S4, July 2019.

[8] Kamatchi Sankaravadivel S, Parthiban Latha. "An effective ways of finding adverse drug reactions in pharmacovigilance". *International Journal of Engineering and Advanced Technology(IJEAT),* ISSN: 2277-3878, Volume-8, Issue-2S4, July 2019.

[9] Kamatchi Sankaravadivel S, Parthiban Latha."Data mining algorithms for pharmacovigilance". *International Journal of Engineering and Advanced Technology (IJEAT),* ISSN:2249-8958, Volume-9, Issue- 1S, October 2019.

Chapter 8

Quantum-Based Deep Learning for Multi-Level Grading of Mangoes

Nirmala Gururaj[1,*], Viji Vinod[1,†] and Latha Parthiban[2,‡]

[1]Computer Applications Department, Dr. M.G.R. Educational and Research Institute, Maduravoyal, Chennai, India

[2]Department of Computer Science, Pondicherry University Community College, Pondicherry, India

Abstract

In today's consumer lifestyle, food habits have become an essential part of individual health. Natural food like fruits and vegetables supplement most of the need for this healthy lifestyle. However, the consumer does not often have access to good-quality fruits or vegetables. The quality of fruits or vegetables is the key to any purchasing decision and improving customer satisfaction. The current grading of fruits based on quality is a primarily visual inspection of physical characteristics like color, size, shape, which are not scalable for the multitude of quality assessments that need to be done. This proposed study analyses the current grading practices of mango fruits to build a fully automated holistic grading assessment, commercially viable and with improved accuracy for bringing quality mangoes from the farm to the customer's plate. This study considers multi-level grading on significant ripening indices of mango based on microscopic, external, and internal attributes using a low-cost portable device and sensor technologies, Deep Learning

* Corresponding Author's Email: gururaj.nirmala@gmail.com.
† Corresponding Author's Email: vijivino@gmail.com.
‡ Corresponding Author's Email: lathaparthiban@yahoo.com.

In: Frontiers in Quantum Computing
Editor: R. Anandan
ISBN: 978-1-68507-816-4
© 2022 Nova Science Publishers, Inc.

Artificial Intelligence, and Spectral Imaging techniques. Such a grading system will help the multiple stakeholders in the mango fruit supply chain deliver quality mangoes to the consumers consistently and cost-effectively.

Keywords: sustainable quality, hyperspectral analysis, near-infrared, total soluble solids, support vector machine, artificial neural network, brix, refractometer

1. Introduction

1.1. Current Grading Methods

India has for long been reputed to be the world's second-largest producer of fruit and vegetables. Nearly 8 to 18% of the fruits and vegetables go wasted during post-harvest handling, according to the report by the APEDA database (2020). One of the main reasons for post-harvest losses is the inefficiencies in the supply chain of perishable commodities. The absence of a consistent grading mechanism makes it challenging to deliver quality products at various places at the right price. Some of the practices for grading still followed today is as listed below

- Manual grading of fruits is done by visual inspection to remove the damaged, diseased and rotten fruits by hand. This process is subjective and prone to human error and is still followed in many countries and rural parts of India. Shown below in Figure 1.1 is an image of manual grading of Alphonso mangoes done by farmworkers into crates at a local mango trader's premises in Ratnagiri, Maharashtra, India.
- Size and color based grading using semi-manual and semi-mechanical methods, For instance, size grading is done using a conveyer belt with varying apertures and then manually weeding out diseased fruits before packaging
- External grading uses appearance-based features like color, shape, weight, skin texture, and specific gravity
- Grading is done only based on internal qualities like firmness and soluble sugars

Figure 1.1. Manual Sorting of mangoes.

All the above processes use only a few parameters and do not consider a holistic multi-level grading process.

1.2. Current Grading Technology

Image processing algorithms with visual inspection techniques improved all fruits and vegetables' manual and semi-manual grading and sorting processes. Some of the technologies most commonly used in grading fruits and vegetables today are listed below

- Computer vision and Image Processing methods are used for appearance-based grading. However, this method cannot do the fine-grain classification of fruit's in-between maturity ripening stages
- Near-Infrared Spectroscopy (NIRS) is a traditional method for internal quality grading. The disadvantage of this technique cannot be done in real-time and requires expensive equipment to measure internal qualities
- Techniques like e-Nose, Magnetic Resonance Imaging, X-ray, Computed Tomography can also do grading based on a few attributes

1.3. Limitations of the Current Grading Methods and Technology

Though image processing algorithms with visual inspection techniques improved the manual grading process, they still considered only a few

parameters like either size or color or size and texture for grading. There is a gap in considering multiple parameters for grading and the process was not holistic, considering both external and internal attributes of the fruit. The image processing technologies also did not help in the fine-grain classification of the ripening maturity stage. Only two stages, raw and ripe were easily identified and the in-between stages of early ripening and partial ripening were challenging to identify. Also, technologies to identify the internal attributes like the soluble sugars require massive lab equipment's like the spectrometer and cannot be used in real-time grading. Most of the other techniques used in quality grading, like the e-Nose and X-ray, though non-destructive can only grade based on a few parameters. These methods need constant tuning to perform better.

1.4. The Need for Artificial Intelligence

Algorithms with more powerful automatic learning capabilities were needed to perfect the grading process. Artificial Intelligence (AI) and Machine Learning algorithms provide the much-needed speed and accuracy to the existing grading process. These methods are fast, reliable, and consistent in their performance. Some of the standard grading processes using AI techniques are

- The features extracted from image processing methods as input to machine learning algorithms for grading and classifying fruits and vegetables. The feature set is based on visual parameters like size, shape, color, texture and the grading accuracy is limited to the feature set selected
- Deep Learning using pre-trained networks for appearance-based grading. This technique is fast and efficient but will require a large dataset and a longer time to train with more computing power

Nandi, Tudu, and Koley (2014) used computer vision methods to grade fruits based on color, size, shape, and texture. The authors used size, maturity level, and fuzzy logic for consistent grading and uniform sorting with an accuracy of around 90% for different grades. Zhenbo et al. (2020) used a pre-trained model VGG-M network for vegetable recognition based on vegetable ages classification with an accuracy of 92.1%. This research aims to design an efficient improved algorithm for holistic grading using artificial intelligence

techniques, digital image processing methods, convolutional neural networks, and spectral imaging techniques. For this study, mango, an abundantly grown fruit in most parts of India, is chosen.

1.5. The Proposed Solution for Fruit Grading

The proposed research study aims to grade mangoes based on appearance and internal quality attributes. The individual objectives of the proposed solution are as follows.

- Grade the fruits using a multi-level grading solution
- Use a combination of image processing, computer vision, and deep learning techniques for external grading with improved accuracy
- The Deep Learning technology used will be a Convolutional Neural Network (CNN) FMG13 model built explicitly for this research on a dataset created exclusively for this study
- Use a combination of machine learning and multivariate algorithms to predict the internal quality attribute of the fruit and classify based on the selected internal attributes
- Use of low-cost sensors and mobile camera/hand-held imaging devices to do the multi-level grading with improved accuracy

Thus, a holistic grading solution is developed to ensure quality mango fruits on the customer's plate.

1.6. Mango Supply and Demand

The acreage records show that Indian Horticulture has grown exponentially over the years and the fruits and vegetable production has surpassed the country's food-grain production. According to the Farm Ministry news (2021), the Mango production in the country is estimated to increase by 4.24% to 21.12 million tonnes in the crop year 2020-21 ending June. Mangoes are grown in almost all states and India occupies one of the top positions in total production among Mango exporters. According to the Agricultural and Processed Food Products Export Development Authority (APEDA), India is home to more than 1000 different mangoes. Around 30 varieties are grown

commercially for export business, according to the report by the APEDA database (2020). Maharashtra is the largest exporter among the various states, followed by Tamil Nadu, which exports up to 14 varieties of mangoes. Banganapalli, Alphonso, Neelam, Rumani, and Bangalore are tropical varieties grown abundantly in Tamil Nadu, catering to the export business. Good quality mangoes are mangoes that consumers will want to purchase again and again. It requires an innovative, intelligent grading solution from the mango supply chain stakeholders to meet the demand for quality mangoes.

1.7. Factors Affecting Mango Quality

The mango maturity at the harvest time is crucial to determine the fruit-eating quality and shelf life (Kader and Yahia 2011). Mango growers always face a challenge as to when to harvest to provide the best quality fruits. Mangoes plucked before optimum maturity will ripen but will not have the desired aroma and flavor. Such mangoes are more susceptible to chilling injury and have a shortened shelf life. Over-mature mangoes become soft and mealy and are insipid when ripe. The selection of mangoes fruits based on maturity are based on several parameters: shape, peel color, peel texture, flesh firmness, internal flesh color development, firmness, and soluble solids content. However, the most commonly used reliable maturity indices are the shoulder development or shoulder fullness above the stem attachment point, the flesh color change from green to yellow, and the cheeks development.

Mango is a climacteric fruit with a short shelf life. Maturity indices not only decide the fruit quality but also need to provide marketing flexibility. Sometimes mangoes will need to be shipped long distances, resulting in the need to harvest them before the ideal maturity stage. So, there is always a compromise between eating quality and market flexibility. Post-harvest handling of mangoes is also an essential factor in deciding the quality of mangoes. Since there is a substantial post-harvest wastage, it is necessary to monitor the ripening quality of the mangoes during the post-harvest handling to ensure quality mangoes to the end customer. The above factors highlight the need for an objective grading of mangoes based on maturity ripening and eating quality.

1.8. Objective Quality Evaluation

The mango quality assessment is done manually in the warehouses, which is high cost, inconsistent due to high subjectivity, and laborious. Quality assessment is typically based on fruit's appearance, texture, flavor, and nutritive value (Kader and Yahia 2011). The farmers will want their mangoes to have a good appearance, few visual defects, be disease-resistant, easy harvest and transportation. For the retailers and wholesalers, the mango fruit's appearance and shelf life are of utmost importance. Consumers not only look at appearance but also the eating quality of the mango fruit for repeated buying. So, various attributes influence the quality grading of the mango fruit (Kader 1999). The objective assessment of mango fruit quality is based on several factors, as listed below.

- Number of parameters to be measured based on the external appearance
- Inclusion of parameters based on the internal properties of the mango fruit
- Mechanisms and techniques used to evaluate the mango fruit in a non-destructive manner
- Correlations between the measured quality to the selected parameters
- Algorithmic analysis of these parameters to the overall consumer quality acceptance

Such an assessment will help farmers, retailers, wholesalers, consumers access quality mango fruits. This process will also help decide which mangoes to export and which to consume locally to minimize post-harvest losses.

1.9. The Significant Contribution of This Research

India is one of the largest exporters of mangoes to various international markets and grading is essential before the mangoes are packaged and sent. Also, it is essential to grade mangoes to understand which mangoes to transport long distances and which ones to consume locally to minimize the post-harvest losses. Post-harvest losses are huge in developing countries due to a lack of knowledge about maturity, harvest time, and management of post-harvest transportation logistics (Prasad, Jacob, and Siddiqui 2017). Current

methods of grading are done manually or semi-manually, are highly inconsistent and are labor-intensive.

To provide a non-destructive quality assessment of mangoes, the significant contribution of this research is

- Providing good quality mangoes to the customer and achieving total customer satisfaction
- Design a holistic multi-level grading of mango based on both external and internal quality attributes with improved performance using the FMG13 CNN model created for this study
- Grade using low-cost affordable portable device technology and using Deep Learning, Machine Learning algorithms, and Multispectral imaging techniques for improved performance

To develop such a system, the following objectives were formulated.

- To grade the mango quality based on microscopic images of the mango skin texture based on the dataset of surface cutin layer images captured using a light microscope clip attached to the mobile phone. Evaluate the mango maturity ripening stage based grading using Convolutional Neural Network (CNN) into one of the Raw, Early Ripe, Partial Ripe, and Ripe stages
- To grade mango quality based on external attributes of the mango. The dataset was created with full mango images captured using a mobile phone camera. The maturity ripening stage was extracted using the Convolutional Neural Network FMG13 model created for this study. Other external attributes like shape, color, texture, and defect features were extracted using computer vision and image processing techniques. Classify the mango quality based on maturity ripening stage, shape, texture, color, and defects into three categories: not fit, Average, and Good quality, comparing five different machine learning algorithms for higher accuracy
- To grade the mango quality based on internal attributes of the mango. Determine the mango sweetness by measuring the Total Soluble Solids (TSS) using Near- Infrared (NIR) multispectral imaging captured by an IoT sensor. Classify the mangoes based on their sweetness index into three classes: low, moderate, and high sweet mango using the Random Forest algorithm (RF)

- Classify the overall quality grade of mangoes based on a combination of the above attributes, resulting in higher accuracy into Grade A, Grade B, and Grade C using the Random Forest algorithm (RF)

2. Literature Survey

New age customers have grown up the value chain and the mango fruit quality rather than price drives their preferences. There is a significant dependency on experts who are trained to pick good quality mangoes. Their absence has led to inconsistent grading, rejections at multiple points resulting in increased wastage, impacting mango quality and pricing. This chapter discusses the various existing methods of grading the quality of mangoes, the challenges, and the gaps. An extensive literature survey was conducted on the following areas to develop an intelligent automatic cost-effective overall grading solution for mangoes.

- Factors influencing the fruit quality assessment
- Assessment based on external appearance using Computer Vision, Image Processing and Convolutional Neural Network methods
- Assessment based on internal attributes using Visible and Near-Infrared spectroscopy
- Quality grading using Classification and Regression algorithms
- Quality grading using different devices

2.1. Factors Influencing the Fruit Quality Assessment

Quality grading of fruits is based on external appearance, internal attributes, nutritional and safety aspects. The external appearance factors include color, size, shape, texture, and defects. The texture is about the skin texture, firmness of the fruit, and its juiciness. The internal attributes include flavors like sweetness, sourness, bitterness, and aroma. Maturity at harvest and post-harvest management determines the fruit quality and also the shelf life of the fruits.

Dhatt et al. (2007) detailed different maturity types in their paper and used visual and physical indices to measure maturity. While describing the visual indices, he said the size and shape of fruit can assess the harvest time maturity.

For example, in mango and stone fruits, the fullness of cheeks adjacent to the pedicel guides the maturity stage.

Table 2.1. Literature survey of standard maturity indices

Maturity indices category	Index	References
Physical	Size	Dhatt et al. (2007)
	Shape	
	Texture	Preethi, Soorianathasundaram, and Subramanian (2016)
	Specific Gravity	Kapse and Katrodia (1997)
	Colour	Li, Lecourt, and Bishop (2018)
Chemical	Soluble Solids Content	Crisosto (1994)
	Oil Content	Furferi, Governi and Volpe (2010)
Physiological	Ethylene Production	Dhall (2013)
Calculated	Elapsed days from full bloom to harvest	Khodabakhshian, Emadi, and Khojastehpour (2017)

Crisosto (1994) determines which maturity indices are consistent in determining the harvested fruit quality. He stated that soluble solids concentration increased with fruit maturity and fruit ripening and an excellent and reliable maturity index. Furferi, Governi, and Volpe (2010) used oil content as the ripening index in olive fruits. Preethi, Soorianathasundaram, and Subramanian (2016), in their study on fruits like mango, observed that the cell wall structure could determine the texture and ripening stage. Below is the summary of some of the indices reviewed in this research in Table 2.1.

2.2. Assessment Based on External Appearance Using Computer Vision, Image Processing and Conventional Neural Network Methods

Image processing methods extract the color RGB/HSI features from images by pre-processing, resizing, removing the background, cropping the images. Kamel (2015) classified tomatoes using different color models like RGB, HSV, CIE L*a*b to create a combined color feature vector input to Artificial Neural Network (ANN) classifier that determines the maturity grading based on color.

Naik and Patel (2017) used computer vision methods to grade fruits based on color, size, shape, and texture. The author used image segmentation

techniques like thresholding and clustering, histogram, and edge detection to analyze the fruit images for meaningful data.

Table 2.2. Literature survey of grading and classification techniques

Method	Fruits analyzed for grading	Index used	References
Image Processing Methods	Tomato	Colour	Kamel (2015)
	Multiple Fruits	Colour, Size, Shape, and Texture	Naik and Patel (2017)
	Tomato	Colour and Texture	Arakeri and Lakshmana (2016)
Convolutional Neural Network	Banana	Ripening Stage	Zhang et al. (2018)
	Mango	Ripening Stage	Yossy et al. (2017)
	Multiple Fruits	Image Features	Israr Hussain, Jianhua He, and Zhu-liang Chen (2018)

Arakeri and Lakshmana (2016) graded tomatoes using image processing methods to detect the diseases and analyze the ripeness. They used color and texture to grade the tomatoes with an accuracy of 96.47%.

2.3. Assessment Based on Internal Attributes Using Visible/Near-Infrared Spectroscopy

Mango eating quality at a fully ripe stage is usually characterized by the mango fruit's Total Soluble Solids or TSS. The starch of the mango fruit is converted to soluble sugar content. Hence, the total carbohydrate content, which measures the Dry Matter (DM) content and soluble sugar content, is a good indicator of the eating quality. Short wavelength Near-Infrared spectroscopy has been widely used to measure the mango internal qualities in a non-destructive manner based on TSS or DM.

Pullanagari and Li (2021) classified the quality of fresh cherry fruits using hyperspectral imaging and multivariate analysis. The authors determined TSS and firmness using Gaussian Process Regression (GPR) to evaluate the prediction performance and predictive uncertainty. Tran and Fukuzawa (2020) classified Kiou apples based on soluble solids and used a portable spectroscope to predict the Soluble Solids Content (SSC). The author used five bands at 535nm, 680nm, 730nm, 760nm, and 900nm and the results were comparable to the traditional dispersive spectrometers.

2.4. Quality Grading Using Classification and Regression Algorithms

Win (2019) classified different mangoes using the Naïve Bayes Algorithm with 94% accuracy. De Goma et al. (2018) identified fruits based on appearance features by comparing KNN, Naïve Bayes, Decision Tree, and Bagging and found KNN to perform better, with an accuracy of 81.9%. Pathanjali et al. (2018) performed food image classification by comparing SVM and KNN and found SVM to perform better. Md Iqbal et al. (2016) used Linear Discriminant Analysis (LDA) to classify citrus fruits based on color. Truong Minh Long and Truong Thinh (2020) compared different machine learning algorithms to grade mangoes based on four external features like length, width, defect, and weight and found Random Forest (RF) to have an accuracy of 98.1%.

2.5. Quality Grading Using Different Devices

Traditionally, bulky computer vision-based systems were used for food grading. Non-destructive evaluation of fruits was done using X-ray imaging, Magnetic Resonance Imaging (MRI), ultrasound, Computed Tomography (CT), and Near-Infrared based imaging. En-Cheng Yang et al. (2006) examined the internal injury in fruits using X-ray. Yacob et al. (2005) compared Magnetic Resonance Imaging (MRI) with X-ray for defect detection in fruits. McGlone et al. (2003) used the traditional lab NIR spectroscopy in the wavelength region of 750-1100nm for internal quality evaluation of mandarin fruit.

3. Materials and Methods

New age customers have grown up the value chain and value freshness and quality of fruits rather than price. There is a significant dependency on the experts to pick the right quality of fruits. The absence of such experts has led to inconsistent grading of fruits and rejections at multiple points resulting in increased wastage. It is necessary to have a uniform way to grade the mangoes and curtail losses during the post-harvest cycle to harness the market potential. This research focuses on the quality grading of mangoes for both internal consumption and export market based on several maturity ripeness indicators

using cost-effective device technology and a combination of different algorithms for higher accuracy.

3.1. Research Objective and Methodology

The main objective of designing an intelligent mango quality grading solution is to establish consistency in grading and increase the accuracy of identifying good quality mangoes in a non-destructive and cost-effective manner. To develop such a system, the following objectives were formulated.

- Grade the mango quality based on microscopic images of the mango
- Grade mango quality based on external attributes of the mango
- Grade the mango quality based on internal attributes of the mango
- Grade based on a combination of attributes, both internal and external

The methodology has the following process and steps that will be carried out to accomplish the grading process in each of the above research objectives

- Data acquisition
- Pre-processing and feature extraction
- Model design

3.2. Data Acquisition

The dataset created for this research is of different types based on the type of grading. For microscopic grading, the mango is kept in the Light-box facing the microscopic lens, and the skin surface image is taken at 100x magnification and using a mobile phone camera to capture full mango images for external grading. The image dataset is created by capturing full mango images of three different varieties Banganapalli (BA), Neelam (NL), and Rumani (RU). For internal grading, the image acquisition device is a NIR imaging system with a NIR sensor chip AS7263 (AMS AG) connected to a Raspberry pi. Shown below in Table 3.1 is the dataset used for the different types of grading.

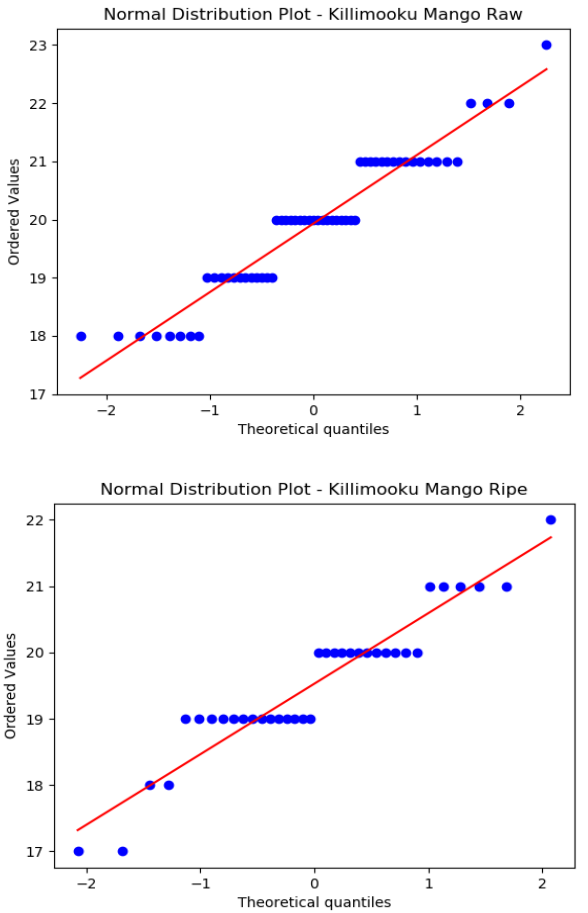


Figure 3.1. Normal distribution plot of skin dataset for Killimooku mangoes.

The image dataset was transformed using Singular Value Decomposition (SVD) according to the cuticle patterns in the texture of the 'Maturity-Stage' skin dataset.

The normalization plot of the SVD values of the dataset shows the dataset to follow a normal distribution and is as shown in Figure 3.1 for Killimooku and Figure 3.2 for the Banganapalli dataset. The results of the D'Agostino's normality tests show (stat = 1.164, p-value = 0.558) for the Killimooku raw mango dataset and (stat = 0.3776, p-value = 0.827) for the Killimooku ripe mango dataset suggesting that the null hypothesis cannot be rejected and the data is normal.

Table 3.1. Dataset used for grading

Type of grading (Dataset name)	Total no in the dataset/dataset type	Mango varieties	Source
Microscopic Grading (Maturity-Stage)	360 skin images	Banganapalli, Killamook (Totapuri)	Tada farm near Chennai
External Grading (Variety-Quality)	1328 full images and external attribute data	Banganapalli, Rumani, Neelam	Tada farm near Chennai
Internal Grading (Ripe-NIR)	1108 spectral attribute data	Banganapalli Rumani, Neelam	Local farms in Chengalpattu.

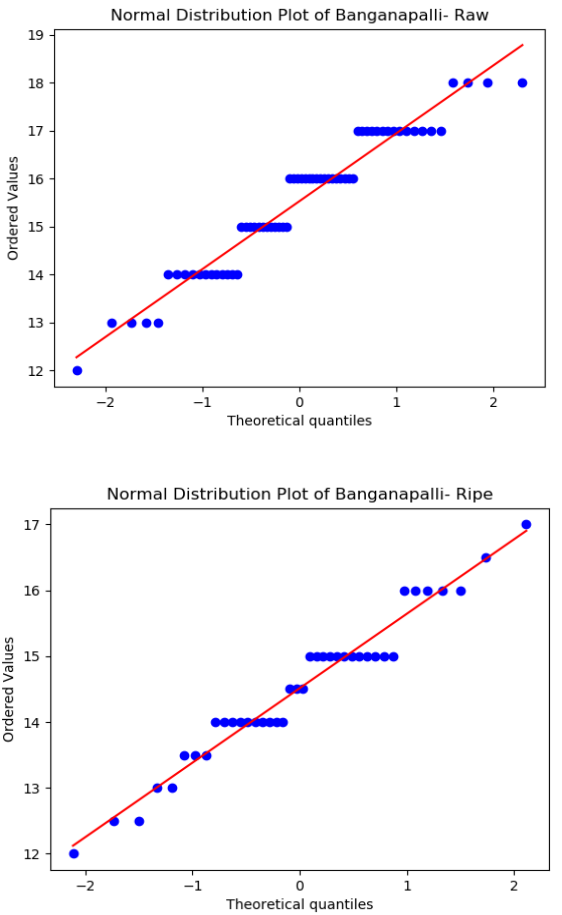


Figure 3.2. Normal distribution plot of skin dataset for Banganapalli mangoes.

The results of the D'Agostino's normality tests show (stat = 2.34, p-value = 0.309) for Banganapalli raw mango dataset and (stat = 0.0335, p-value = 0.983) for Banganapalli ripe mango dataset suggesting that the null hypothesis cannot be rejected and the data is normal at 95% confidence level. The attribute dataset used for external grading is normalized to bring the columns of the attribute dataset to a standard scale without distorting differences in the ranges of values and to make the model perform better. The standard scalar function is used for the normalization of the dataset using the standardization technique. This function uses the arithmetic mean and the standard deviation for normalizing the data. The normalization plot of the external attributes of three varieties of mango of the 'Variety-Quality' dataset is as shown below in Figure 3.3.

The results of the D'Agostino's normality tests shows (stat = 3.0609, p-value = 0.216) for Rumani, (stat = 2.467, p-value = 0.2911) for Neelam and (stat = 4.80, p value = 0.090) for Banganapalli dataset suggesting that the null hypothesis cannot be rejected and the data is normal at 95% confidence level. A one-way Analysis of Variance (ANOVA) was performed on the 'Variety-Quality 'dataset and results as shown in Table 3.2.

A significant difference was observed in the variety of the mangoes concerning color, size, shape, and appearance as the p values of all these attributes are less than 0.05. A two-way ANOVA of the size and appearance attribute was done to understand the interaction between these attributes concerning variety and the results are shown in Table 3.3.

Table 3.2. Differences in attributes influencing variety recognition - one-way ANOVA

External attributes	Type of analysis	Sum of squares	Df	F	Sig.
Size	Between Groups	248..97	2.0	315.13	$4.02\ e^{-110}$
	Within Groups	460.21	1165		
Colour	Between Groups	0.3503	2.0	133.19	$8.07\ e^{-53}$
	Within Groups	1.532	1165		
Shape	Between Groups	330.13	2.0	6.120	0.0022
	Within Groups	31417.12	1165		
Appearance	Between Groups	2560.70	2.0	21.676	$5.715e^{-10}$
	Within Groups	68813.36	1165		

The inference of the two-way ANOVA is that the attribute's size and appearance have p values less than 0.05, this means both these attributes have a significant effect on the variety. And the p-value of their interaction is also less than 0.05 indicating significant interaction between them. A one-way

Analysis of Variance (ANOVA) was performed on the 'Ripe-NIR' dataset and a significant difference was observed in the variety of the mangoes concerning TSS value ($F(2, 972) = 2460.64$, $p<0.05$) as shown in Table 3.4.

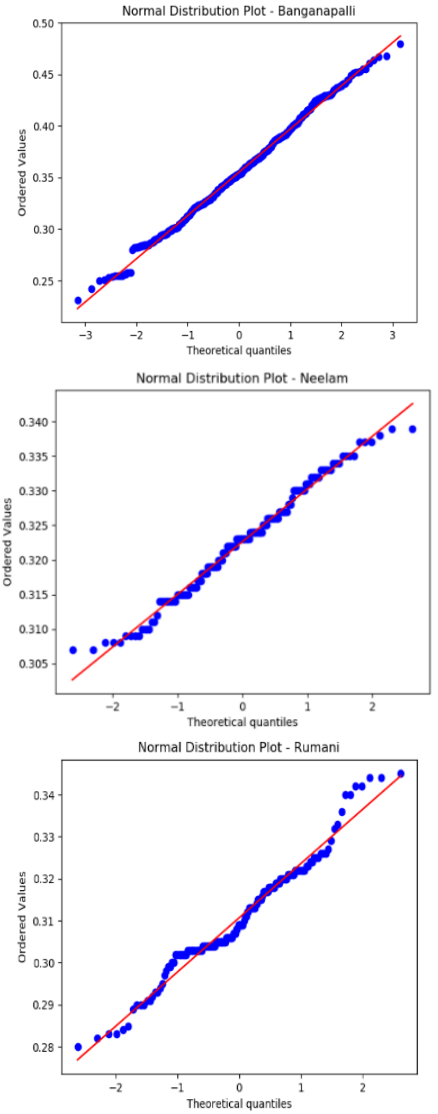


Figure 3.3. Normal distribution plot of Variety-Quality dataset.

Table 3.3. Interaction between size and appearance - two-way ANOVA

External attributes	Sum of squares	df	F	Sig.
C(Variety)	0.0897	2	44.37	$2.67e^{-19}$
Appearance	0.0135	1	13.38	$2.64e^{-04}$
Size: Appearance	0.3420	1	338.04	$1.79e^{-66}$
Within Groups	1.1767	1163		

The normalization plot of the TSS value of the Banganapalli, Neelam, and Rumani varieties of mango of the 'Ripe-NIR' dataset is as shown above in Figure 3.4.

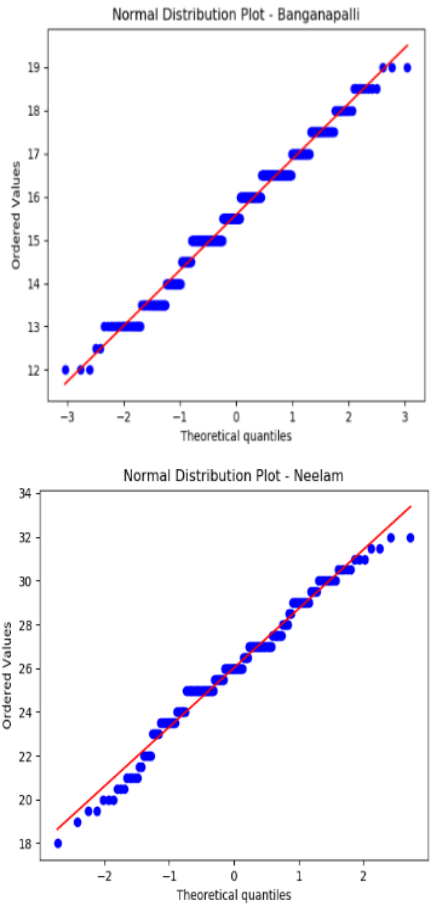


Figure 3.4. (Continued).

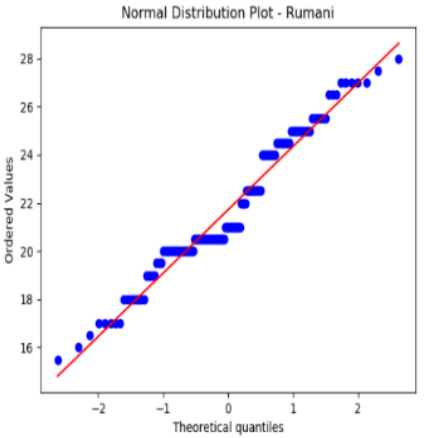


Figure 3.4. Normal distribution plot of the TSS dataset.

Table 3.4. Differences in internal attributes that influence a variety of one-way ANOVA

External attributes	Type of analysis	Sum of squares	df	F	Sig.
TSS	Between Groups	18772.88	2.0	2460.64	$5.57e^{-27}$
	Within Groups	3707.82	972		

The results of the D'Agostino's normality tests shows (stat = 2.784, p value = 0.2485) for Banganapalli dataset, (stat = 3.563, p value = 0.168) for Neelam dataset and (stat = 4.725, p value = 0.0941) for Rumani dataset suggesting that the null hypothesis cannot be rejected and the data is normal at 95% confidence interval.

3.3. Process and Steps

3.3.1. Research Objective 1: Microscopic Grading

The first research objective is to grade mangoes based on the microscopic images and arrive at the mango quality by determining the maturity stage based on the cutin layers and reticulate patterns visible on the mango peel surface. By analyzing the surface texture properties, subtle differences can be identified with reasonable accuracy in the various ripening stages. The mango texture has a layer of wax and cuticles on the surface that changes during fruit development.

Pre-Processing and Feature Extraction

A grayscale transformation of the microscopic surface images was done to convert the RGB image to a grayscale image. The image noise was then reduced by applying the Gaussian blur.

Model Design

In this research, the Convolutional Neural Network (CNN) model will correlate the mango skin texture to its maturity ripening stage. The model will classify the mango's skin structure into one of the four stages: Raw, Early Ripe, Partial Ripe, and Ripe. The mango quality is based on these four stages and only the ripe mango is of eating quality. The model is built using Keras and TensorFlow libraries in python 3.7.

3.3.2. Research Objective 2: External Grading

The second research objective is to grade mangoes based on external attributes. The image dataset was created with full mango images captured using a mobile phone camera. A total of 15 attributes were extracted for grading and classifying the mango according to its variety. The attributes extracted are mango ripening stage, color, shape-based features like mango roundness, aspect ratio, area, perimeter, size, width, and length, texture-based features like homogeneity, contrast, entropy, energy, correlation, dissimilarity, and defects.

Pre-Processing and Feature Extraction

The images are converted to grayscale from RGB and then to reduce the noise, a Gaussian blur filter is applied. Using the canny edge detection method, the vertices of the mango are determined.

Model Design

The model is based on machine learning classification algorithms that will take the mango external features extracted above (maturity ripening stage, color, shape-based, and texture-based features) as input and classify and grade the mango. Based on the 15 input features, the model can classify the mango as Banganapalli, Neelam, and Rumani and grade the mango quality as NotFit, Average and Good.

The model was designed using five different classifiers in succession to identify the best-suited classifier with high accuracy. The architecture diagram of the model is as shown below in Figure 3.5.

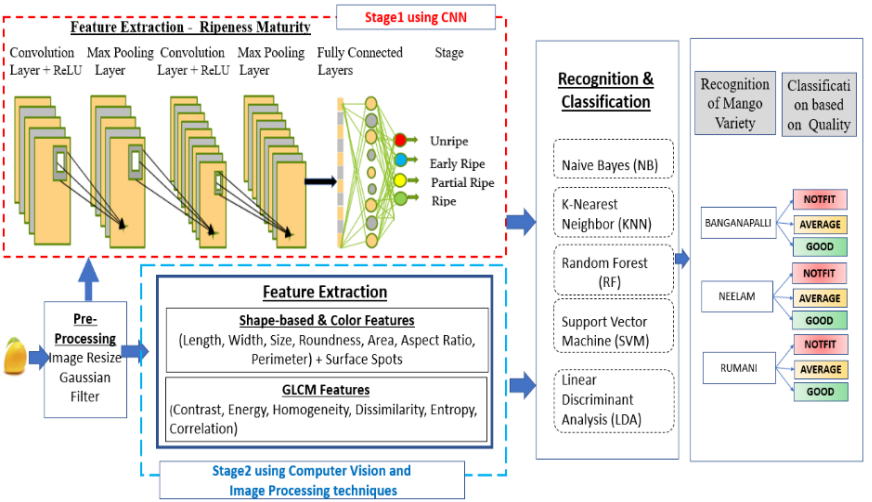


Figure 3.5. Model design based on external attributes of the mango.

3.3.3. Research Objective 3: Internal Grading

The third research objective is to grade mangoes based on internal attributes. Total Soluble Solids (TSS) is one such internal attribute used to indicate the mango sweetness level. TSS is measured destructively using Brix Refractometer and the value of TSS is represented in Brix degrees. The TSS is measured using a cost-effective portable sensor and a non-destructive technique like Near-Infrared (NIR) multispectral analysis.

Pre-Processing and Feature Extraction

The spectral data is pre-processed to eliminate the noise and scattering effects using scatter correction methods or spectral derivatives methods. In this research study, the spectral derivative Savitzky-Golay (S-G) polynomial filters reduce the spectral signal-to-noise ratio. The NIR spectral reflectance and absorbance data are pre-treated with the Savitzky-Golay filter with a three-point window and second-order polynomial.

Model Design

The model is based on multivariate regression techniques to get meaningful insights into the spectral data correlation to the internal quality parameters. The model will analyze the spectral readings and predict the TSS of the mango. Based on the TSS value, the mango will be graded as High Sweet, Low Sweet, and Average Sweet mango. The model used two non-parametric

regression techniques, Support Vector Machine Regression (SVMR) and Artificial Neural Network (ANN). Both were analyzed to identify the best-suited model with higher accuracy in predicting the TSS value. Random Forest classifier was then used to grade mangoes based on sweetness level.

3.3.4. Research Objective 4: Combined Multi-Level Grading

This research aims to build an intelligent automatic overall quality grading solution by combining the appearance and internal flavors. The solution is aimed at selecting the appropriate maturity indices both subjectively and objectively. Subjective indices are measured qualitatively and are related to color, shape, and texture-based features. Objective indices are measurable quantitative indices relating to the TSS of the mango fruit.

4. Data Analysis and Findings

This chapter discusses the model development and execution process flow of the grading methodology discussed in the previous section to achieve the different research objectives. The model development will discuss the nature of the mango dataset composition used and the dataset split recommended for training and testing the model. The execution process flow analyses the model results and the different techniques used for classifying and grading the mangoes. The model development and execution output are compared with already established works to highlight improvement and efficiency in the current study.

4.1. Step 1: Creation of Dataset

The ‘Maturity-Stage' dataset is created using microscopic images of the mango surface cutin texture of two varieties: Banganapalli and Killimooku. The images were captured at four different maturity stages: Raw, Early Ripe, Partial Ripe, and Ripe. The dataset was split as 60:20:20 with 60% of mango images used for training the model, 20% for testing the model, 20% for validating the model.

The 'Ripe-Stage' dataset is created using full mango images captured by the mobile phone camera and has 12,168 mango images across four different maturity ripening stages: Raw, Early Partial Ripe, and Ripe. The model based

on Convolutional Neural Network is trained on the 'Ripe-Stage' dataset. The dataset split used here is 60:20:20.

The 'Variety-Quality' dataset consists of 1328 mango fruit images across three mango varieties Banganapalli, Neelam, Rumani. The dataset is used for feature extraction of stage, color, shape, texture, defects. The 'Ripe-NIR' dataset of 1108 multispectral data readings of 360 mangoes across the same varieties was used to predict TSS. The pre-processed spectral data was divided into two subsets, where 60% of the data was used for training and 40% was used for testing and validating the model.

4.1.1. Model Development and Execution Process Flow

Microscopic Grading

Research Objective 1 is to grade mango quality by determining the maturity ripening stage using microscopic mango images of the surface skin texture. The model is trained on the grayscale transformation of the microscopic surface images on the 'Maturity-Stage' dataset. The image noise was then reduced by applying the Gaussian blur. The model prediction of the maturity stage is by analyzing the cutin texture visible at 100x magnification of the BEILESHI clip-on phone microscopic images.

External Grading

Research Objective 2 is to grade mangoes based on the mango external attributes captured by the mobile phone camera. In this study, Convolutional Neural Network (CNN) FMG13 model is used as a feature extractor to get the maturity ripening stage. The other external attributes, like color, shape, texture features, and defects, are extracted using Computer vision and Image processing techniques. The combined set is used for mango external quality grading and variety recognition. The steps in model development and execution process flow are shown in Figure 4.1 below.

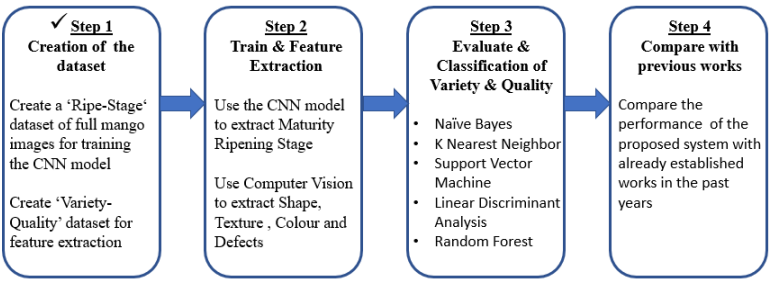


Figure 4.1. Process flow for grading based on external attributes.

4.2. Step 2: Train the Model

The FMG13 model is trained on the 'Ripe-Stage' to predict the maturity ripening stage of the mango. Once trained, the network accuracy with the test dataset was 93.4%, with a loss of 3.3%. The shape-based features area, perimeter, major-axis length, minor-axis length, roundness ratio, aspect ratio, and defects were extracted using computer vision and image processing methods from the 'Variety-Quality' dataset. The CIELab model helps to identify the different colors of the mango according to its ripeness stage. The texture features represent how smooth or coarse the mango surface is and help detect any surface defects.

4.3. Step 3: Evaluate and Classification of Variety and Quality

The features extracted are input to the classifier to recognize the mango variety and grade the mango quality as NotFit, Average and Good. Their performance of various classifiers was evaluated to determine the best-suited classifier for mango variety recognition and quality grade classification. The chart below in Figure 4.2 shows the mango variety grading and quality grading and their average grading accuracy.

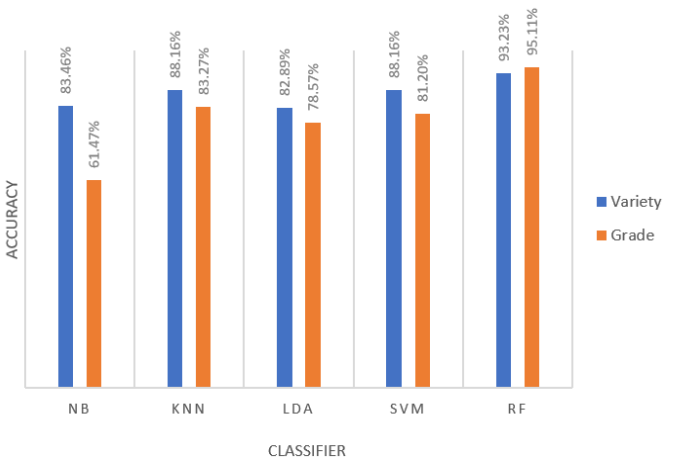


Figure 4.2. Comparison of different classifiers for variety and quality grading.

4.3.1. Step 1: Model Development and Execution Process Flow – Internal Grading

Objective 3 is to grade mangoes based on the internal attributes of the mango. The multispectral chipset of the NIR system is used to determine the quality of sweetness of the mango by predicting the Total Soluble Sugars (TSS). The multivariate model-based classifier will classify mangoes into high, moderate, or low sweet mangoes based on their TSS value.

4.3.2. Step 2: Modelling Using Multivariate Algorithms

Modeling using multivariate algorithms was done on the 'Ripe-NIR' dataset. The reflectance spectra were captured using the NIR sensor chip AS7632 in the wavelength region of 610nm to 860nm. The spectral data was treated with several pre-processing methods of Savitzky-Golay smoothing (S-G) first (1-Deriv) and second (2-Deriv) derivatives, Standard Normal Variate (SNV), Multiplicative Scatter Correction (MSC). Principal Component Analysis (PCA) was applied to the spectral data to analyze sweetness clusters based on the variety seen. The spectral data was pre-processed to smoothen and remove all the noise. Several multivariate models were applied to the pre-processed data to determine the best-suited model for accurate TSS prediction. Models which can overcome multicollinearity and analyze the noisy and redundant variables were considered. Then regression techniques used to predict the TSS of mango are SVM Regression (SVMR) and Artificial Neural Network (ANN). The model performance for TSS prediction of the testing dataset using ANN is shown below in Figure 4.3.

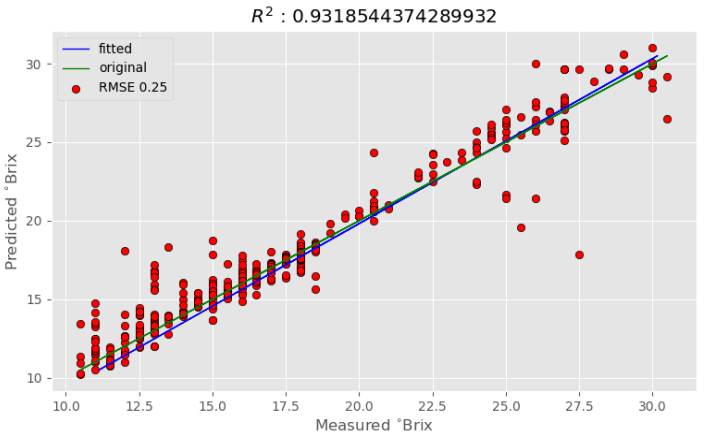


Figure 4.3. ANN Prediction vs Measured Brix values for the testing set.

4.3.3. Step 3: Evaluate and Classify Based on Sweetness/TSS

Once the TSS value was determined, the Random Forest classifier was used to classify the mangoes according to their sweetness value.

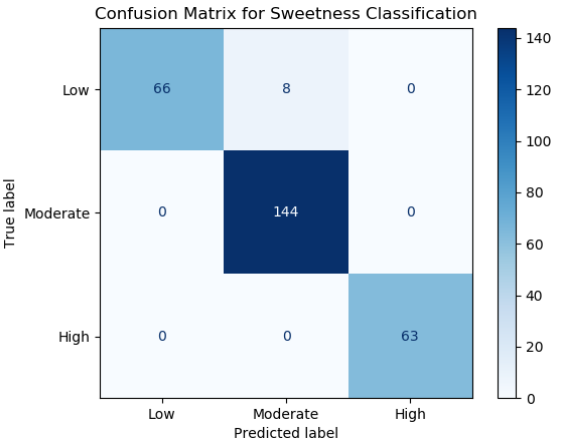


Figure 4.4. Confusion matrix for sweetness classification.

The overall accuracy of the classification was 97.1%. The confusion matrix using the Random Forest classifier is as shown above in Figure 4.4.

4.4. Step 4: Compare with Previous Works

The performance of all the three different types of grading was compared with previously established works. It was found that the proposed solution using the combination of algorithms on the dataset created for this study is better than the performance of similar work done in the past years. Random Forest classifier has the highest accuracy in grade classification and variety recognition based on external attributes. The accuracy for grade classification is at 95.11% and variety is at 93.23%. A detailed comparison of the three different grading types with previous works is presented in the next section.

4.4.1. Model Development and Execution Process Flow – Combined Multi-Level Grading

Research Objective 4 is to classify the overall quality grade of mangoes based on a combination of attributes and methods that results in higher accuracy. The model will perform multi-level grading and evaluate the overall quality

grade based on external appearance attributes like color, shape, texture, defects, and internal attributes such as Total Soluble Sugars (TSS). The mangoes are classified into three different grades: Grade A, Grade B, and Grade C. Grade A is of the highest quality, Grade B is of average quality and Grade C is below average quality.

5. Discussion of the Findings

This section discusses an intelligent overall quality grading solution based on mango appearance and internal flavor. Such a system will use affordable low-cost portable devices for data acquisition that can be taken to the field to perform the grading. Based on the previous three objectives data analysis and findings, an intelligent grading solution based on a combined approach will be built to grade the mangoes into three classes Grade A, Grade B, and Grade C.

5.1. Research Objective

After data analysis and findings of the first three objectives, it was found that grading based on microscopic images considers only one attribute, which is the maturity ripening stage for determining the quality. The surface cutin texture analysis to determine the maturity stage was an elaborate process that requires good precision in capturing the images. This approach resulted in 83% accuracy only and hence not a viable approach for quality grading. The proposed solution considered objective 2 and objective 3 for building an intelligent overall quality grading solution. Such a solution will ensure utmost customer satisfaction and replace the manual, laborious sorting and grading process. The solution will also minimize the post-harvest losses and standardize the quality grading for both internal and export markets.

5.2. Research Methodology for Multi-Level Grading

The proposed solution uses computer vision and image processing techniques for external feature extraction and grades the mangoes as NotFit, Average, and Good Quality. Convolutional Neural Network determines the mango maturity ripening stage as Raw, Early Ripe, Partial Ripe, and Ripe. Near-Infrared (NIR) spectral data analysis determines the mango TSS value. This value is input to

the Random Forest classifier to grade the mangoes into High, Moderate, and Low Sweet. The overall grading of mango quality into three classes Grade A being the best quality mangoes, Grade B of average quality, and Grade C of below-average quality, is determined based on grading rules.

5.3. Data Acquisition

The image dataset created with full mango images of three different varieties Banganapalli (BA), Neelam (NL), and Rumani (RU) was captured using a mobile phone camera. A total of 15 external attributes, maturity ripening stage, color, shape, texture, and defects were extracted from the mobile phone camera images. The sweetness attribute, a measure of total soluble solids, was extracted using Near-Infrared spectroscopy sensor chip AS7263 (AMS AG) connected to a Raspberry pi. The final multi-level grading is done, considering all the external attributes and the sweetness of the mango.

5.4. Pre-Processing and Feature Extraction

The full mango mobile phone images are initially converted to grayscale from RGB and then to reduce the noise, a Gaussian blur filter is applied. Using the canny edge detection method, the vertices of the mango are determined. Then the mango image is segmented from the background and the shape-based features are extracted. The Convolutional Neural Network (CNN) FMG13 model is used to determine the mango maturity ripening stage and the color of the mango is determined using the CIELab coordinates. The texture features are extracted using the Gray Level Co-occurrence Matrix (GLCM). The AS7263 (AMS AG) sensor chip spectral data is pre-processed using Savitzky-Golay filters to remove the noise. Multivariate modeling using an Artificial Neural Network was used to predict the TSS value.

5.5. Model Design

The model is based on grading rules that will take the mango external and internal features extracted above as input and grade the mango as Grade A, best quality, Grade B, average, and Grade C, and below-average quality using the Random Forest algorithm. Random Forest algorithm is used for the

classification of variety, external quality, and internal sweetness grading. Convolutional Neural Network FMG13 model was used to determine the maturity ripening stage of the mango. Artificial Neural Network Regression was used to determine the internal TSS Value.

5.6. Model Development and Execution Process Flow

A dataset of 360 mangoes across three different varieties: Banganapalli, Neelam, and Rumani was used to get the external attributes and the internal TSS value. Random Forest algorithm was used to determine grade based on external qualities and sweetness based on TSS value. The model performance was also analyzed using metrics like Precision, Recall, and F1 Score and shown in Table 5.1. The AUC curve is above 0.99 for all grade stages, indicating that the model can accurately identify the different grade stages.

Table 5.1. Confusion matrix metrics

Category	Precision (%)	Recall (%)	F1 Score (%)	AUC
Grade A	93%	83%	87%	0.9986
Grade B	98%	98%	98%	0.9962
Grade C	98%	99%	98%	0.9984

5.7. Compare with Previous Works

The proposed solution of multi-level grading using the combined algorithm was compared with other similar works established recently. The results are summarised in Table 5.2.

Summary and Conclusion

The traditional mango quality assessment and grading were done manually and semi-manually by farm experts and warehouses. The grading process was highly inconsistent due to high subjectivity and was labor-intensive, proving very costly. The literature survey reflects a gap in the multi-level grading of mangoes based on external and internal attributes. There is also a gap in cost-effective portable devices that can be used for field grading. An improvement in the accuracy of the grading algorithms was also essential.

To ensure quality fruits to consumers' needs, a consistent and affordable, and efficient grading process. This research aims to develop an intelligent mango grading system to automatically grade the overall mango quality based on external and internal attributes using low-cost affordable portable device technology with improved performance. The proposed solution has a novel hardware and software component included for the overall grading of mangoes. The hardware component uses a pre-calibrated multispectral chipset (AS7263 sensor) to capture spectral data and a mobile phone camera to capture full mango images.

The software uses a non-destructive technique like Near-Infrared (NIR) multispectral analysis across six bands in the wavelength region of 610nm to 860nm of the spectral data to extract the TSS values. Computer vision and Image processing techniques were used to extract external attributes and Convolutional Neural Network extracted the maturity ripening stage from full mango images. The external attributes considered are shape-based features, texture-based features, and defect details. The maturity ripening stage classifies mangoes as Raw, Early Ripe, Partial Ripe, and Ripe. The internal sweetness of the mango fruit is determined by measuring the TSS value of the mango fruit.

The external attributes, maturity ripening stage, and the TSS value are input to the Random Forest algorithm for an overall grading of mangoes into three classes Grade A being the best quality mangoes, Grade B of average quality, and Grade C of below-average quality. Implications of the findings. Grading using external appearance features can be done quickly and in real-time. The accuracy of this external grading is increased in this study by doing fine-grain classification of the ripening maturity stage of the mango using the convolutional neural network FMG13 model created for this study to improve the overall accuracy of grading. The total number of features used for grading is 15, a sizeable feature set compared to the existing works done in grading fruits. The dataset used in this study is a uniquely created labeled dataset and reasonably more extensive than the dataset used in the previous works. Both the above process led to more accurate predictions and almost in real-time. The internal grading of the mango fruit determined using a sensor chip also can be implemented in real-time to grade the quality of the mango fruit.

Table 5.2. Comparison with previous research works for external grading

Reference	Sample type	No of classes	Accuracy	Comparison feature
Nandi, Tudu, and Koley (2014)	Five mango varieties (Kumrapali, Amrapali, Sori, Langra, Himsagar)	4 Ripening Stages (Raw, Semi-matured, Matured, Over-matured)	88% - 90%	Size, maturity level grading using Gaussian Mixture Model and Fuzzy logic
Supekar and Wakode (2020)	Dasher and Kesar mango	4 grades (Grade1, 2, 3, and 4)	88.88%	Ripeness, size, shape, and defect-based grading using Image Processing and grading using manual formula
Salunkhe and Patil (2015)	Alphonso mango	4 Ripening Stages (Green, Unripen, Ripen, Fully ripen)	90.4% (RGB) 84.2%(HSV)	Rule-based classification of ripening stage grading using RGB and HSV color model
Naik Sapan (2019)	Kesar mango	4 grades (Class I, Class II, Class III, and Class IV)	83.97%	Shape, size, and maturity parameter-based grading using pre-trained CNN model MobileNet and SVM classifier
Tran and Fukazawa (2020)	80 Kiou apples	TSS Value	86%	Pre-calibrated Multispectral Sensor Chipset based TSS prediction
Nguyen et al. (2020)	Cat Hoa Loc mangoes	2 Grades (Low Sweet, High Sweet)	82%	Sweetness grading of mangoes using VIS-NIR multispectral sensors and classification by Random Forest.

Table 5.2. (Continued)

Reference	Sample type	No of classes	Accuracy	Comparison feature
Proposed Solution	Banganapalli, Neelam, Rumani. 12,168 images 1328 images 1328 images	4 Ripening Stages (Raw, Early Ripe, Partial Ripe, Ripe) 3 Varieties 3 Grades (NotFit, Average, Good) TSS Value	93.4% 93.23% 95.11% 92.3%	Ripening maturity stage grading using CNN FMG13 model Appearance-based recognition for a variety Appearance-based quality grading based on 15 attributes Total Soluble Solids prediction using ANN
	1108 spectral readings 1108 spectral readings 360 mangoes	3 Grades (Low, Moderate and High sweet) 3 Grades (Grade A, Grade B, Grade C)	97.1% 97.5%	Sweetness Classification Holistic multi-level grading based on external and internal attributes

Contribution of the Current Work

The system was validated for three different types of mangoes Rumani, Neelam, and Banganapalli. The current study can recognize the three varieties given the feature set. The variety recognition is in addition to grading the mango fruit based on external appearance features into three grades: NotFit, Average and Good. The study can also predict the mango fruit's TSS (Total Soluble Sugar) in a non-destructive manner which can help decide the eating quality. The TSS directly correlates to the sweetness of the mango and the mango has been graded into Low, Moderate, and High sweet based on the different TSS threshold levels. This process of grading based on sweetness is a unique contribution of this study. The study also compared five different machine learning algorithms for classification and found Random Forest to perform better in fruit variety recognition and external, internal grading.

The final quality grading of the mango fruit is based on a holistic multi-level grading of the mango combining the external and internal grading methods. Such a grading can be done in real-time using low-cost technology and improved accuracy using Deep Learning and Machine Learning algorithms.

Limitations

This study is limited only to three varieties of mangoes which are grown abundantly in South India. With the change in geography, soil, and climate, the TSS value of the mango fruit will vary, and hence the sweetness grading process will also need recalibration. The study does not consider overripe mangoes for grading. Data collection of mangoes across different geographies was a challenge, so this study was limited to mangoes grown in farms around Chennai for grading. For external grading, the illumination needs to be constant and can impact the feature set extracted. The soil, climate can also affect the external features like size and shape and will need to consider a wider pool of mangoes from different geographies. The study currently limits the grading based on defects on the external surface of the mango fruit and does not consider internal defects or infected worm mangoes for quality grading.

Recommendations for Future

The overall feature set can also be extended to include additional internal attributes like Dry Matter (DM), firmness for grading. The sensor-based spectral analysis accuracy can be improved by calibrating it across a wider wavelength region and including more parameters that can be correlated. The solution can be extended to other varieties of mangoes and other climacteric fruits.

References

Arakeri, MP & Lakshmana 2016, 'Computer Vision-Based Fruit Grading System for Quality Evaluation of Tomato in Agriculture industry', *Procedia Computer Science*, vol. 79, pp. 426-433.

Crisosto, CH 1994, 'Stone fruit maturity indices: a descriptive review', *Postharvest News and Information Journal*, vol. 5, no. 6, pp. 65-96.

De Goma, JC, Quilas, CAM, Valerio, MAB, Young, JJP & Sauli, Z 2018, 'Fruit recognition using surface and geometric information, *Journal of Telecommunication, Electronic and Computer Engineering*, vol. 10, no. 1-15, pp. 39-42.

Dhall, RK 2013, 'Ethylene in post-harvest quality management of horticultural crops: A review', *A Journal of Crop Science and Technology*, vol. 2, no. 2, pp. 9-25.

Dhatt, AS, Mahajan, BVC, Sandhu, KS, Garg, A & Sharma, SR 2007, *Handbook on Postharvest Handling of Fruits and Vegetables*, 3rd edition, Punjab Horticultural Postharvest Technology Centre, Punjab Agricultural University, Ludhiana, pp. 1-30.

En-Cheng Yang, Man-Miao Yang, Ling-Hsiu Liao & Wen-Yen Wu 2006, 'Non-Destructive Quarantine Technique- Potential Application of Using X-ray Images to Detect Early Infestations Caused by Oriental Fruit Fly (Bactrocera dorsalis) (Diptera: Tephritidae) in Fruit', *Formosan Entomology*, vol. 26, pp. 171-186.

Furferi, R, Governi, L & Volpe, Y 2010, 'ANN-based method for olive Ripening Index automatic prediction', *Journal of Food Engineering*, vol. 101, no. 3, pp. 318-328.

Israr Hussain, Qianhua He & Zhu-Liang Chen 2018, 'Automatic Fruit Recognition Based on DCNN for Commercial Source Trace System', *International Journal of Computer Science Applications*, vol. 8, pp. 01-14.

Kader, AA 1999, 'Fruit maturity, ripening and quality relationships', *ISHS Acta Horticulturae*, vol. 485, pp. 203-208.

Kader, AA & Yahia, EM 2011, '*Postharvest biology of tropical and subtropical fruits*, Woodhead Publishing Limited, Cambridge, 2011.

Kamel, RM 2015, 'A Study on Colour Sorting of Tomatoes Maturity Using Machine Vision and Artificial Neural Networks', *Egyptian Journal of Agricultural Research*, vol. 93, no. 1, pp. 147-161.

Kapse, BM & Katrodia, JS 1997, 'Ripening behavior of Kesar mangoes to specific gravity, *Acta Horticulturae*, vol. 455, pp. 669-678.

Khodabakhshian, R, Emadi, B & Khojastehpour, M 2017 'Determining quality and maturity of pomegranates using multispectral imaging, *Journal of the Saudi Society of Agricultural Sciences.* King Saud University, vol. 16, no. 4, pp. 322-331.

Li, B, Lecourt, J & Bishop, G 2018, 'Advances in Non-Destructive Early Assessment of Fruit Ripeness towards Defining Optimal Time of Harvest and Yield Prediction—A Review', *Plants*, vol. 7, no. 1, ISSN 2223-7747.

McGlone, VA, Fraser, DG, Jordan, RB & Kunnemeyer, R 2003, 'Internal quality assessment of mandarin fruit by VIS/NIR spectroscopy, *Journal of Near Infrared Spectroscopy*, vol. 11, pp. 323-332.

Md Iqbal, S, Gopal, A, Sankaranarayanan, PE & Nair Athira, B 2016, 'Classification of Selected Citrus Fruits Based on Colour Using Machine Vision System', *International Journal of Food Properties*, vol. 19, pp. 272-288.

Naik Sapan 2019, in *Proceedings of International Conference on Sustainable Computing in Science, Technology and Management, Amity University Rajasthan, Jaipur - India, February 26-28, 2019*: Non-Destructive Mango (Mangifera Indica L., CV. Kesar) Grading Using Convolutional Neural Network and Support Vector Machine, Rajasthan.

Naik, S & Patel, B 2017, 'Machine Vision-based Fruit Classification and Grading-A Review', *International Journal of Computer Applications*, vol. 170, no. 9, pp. 22-34.

Nandi, CS, Tudu, B & Koley, C 2014, 'Machine Vision-Based Techniques for Automatic Mango Fruit Sorting and Grading Based on Maturity Level and Size', *Journal of Sensing Technology: Current Status and Future Trends II*, vol. 8, pp. 27-46.

Nguyen, C-N, Phan, Q-T, Tran, N-T, Fukuzawa, M, Nguyen, P-L & Nguyen, C-N 2020, 'Precise Sweetness Grading of Mangoes (Mangifera indica L.) Based on Random Forest Technique with Low-Cost Multispectral Sensors', *IEEE Access*, vol. 8, pp. 212371-212382.

Pathanjali, C, Vimuktha, ES, Jalaja, G & Latha, 2018, 'A comparative study of Indian food image classification using k-NN and SVMs', *International Journal of Engineering and Technology*, vol. 7, no. 3.12, pp. 521-525.

Prasad, K, Jacob, S & Siddiqui, MW 2017, '*Fruit Maturity, Harvesting and Quality Standards*', *Preharvest Modulation of Postharvest Fruit and Vegetable Quality*, Academic Press, ed. by Siddiqui MW, Academic Press, London, pp. 41-69.

Preethi, P, Soorianathasundaram, K & Subramanian, KS 2016, 'Studies on ultrastructural changes during mango fruit ripening cv. Bangalore, *Biochemical and Cellular Archives*, vol. 14, no. 2, pp. 391-393.

Pullanagari, RR & Li, M 2021, 'Uncertainty assessment for firmness and total soluble solids of sweet cherries using hyperspectral imaging and multivariate statistics', *Journal of Food Engineering*, vol. 289, ISSN 0260-8774.

Salunkhe, RP & Patil, AA 2015, in *Proceedings of the 2015 3rd International Conference on Image Information Processing, Waknaghat, December 2015*: Image processing for mango ripening stage detection, RGB and HSV method, pp. 362-365, New Jersey, USA.

Supekar, AD, & Wakode, M 2020, 'Multi-Parameter Based Mango Grading Using Image Processing and Machine Learning Techniques', *INFOCOMP Journal of Computer Science*, vol. 19, no. 2, pp. 175-187.

Tran, N-T & Fukuzawa, M 2020, 'A Portable Spectrometric System for Quantitative Prediction of the Soluble Solids Content of Apples with a pre-calibrated multispectral sensor chipset', *Sensors*, vol. 20, no. 20, pp. 1-11.

Truong Minh Long, N & Truong Thinh, N 2020, 'Using Machine Learning to Grade the Mango's Quality Based on External Features Captured by Vision System', *Applied Science 2020*, vol. 10, no. 17, ISSN 2076-3417.

Win, O 2019, 'Classification of Mango Fruit Varieties using Naive Bayes Algorithm', *International Journal of Trend in Scientific Research and Development*, vol. 3, no. 5, pp. 1475-1478.

Yacob, Yasmin, Hasnath, Ahmad, Putesh, Saad, Rafikha, Aliana A, Roaf & Sabarina Ismail 2005, in *Proceedings of the International Conference on Information Technology and Multimedia*, Uniten, Malaysia 2005: A comparison between X-ray and MRI in post-harvest Non-Destructive detection method, Malaysia.

Yossy, Emny, Pranata, Jhonny, Wijaya, Tommy, Hermawan, Heri & Budiharto, Widodo 2017, 'Fruit Sortation System using Neural Network and Computer Vision, *Procedia Computer Science*, vol. 116. pp. 596-603.

Zhang, Y, Lian, J, Fan, M & Zhang, Y 2018, 'Deep indicator for fine-grained classification of banana ripening stages', *EURASIP Journal Image Video Process*, vol. 2018, no. 1, p. 46.

Zhenbo, Li, Fei, Li, Ling, Zhu & Jun Yue 2020, 'Vegetable Recognition and Classification Based on Improved VGG Deep Learning Network Model', *International Journal of Computational Intelligence Systems*, vol. 13, no. 1, pp. 559-564.

Chapter 9

Efficient Quantum-Based Secure Route Creation and Data Transfer in Mobile Ad-Hoc Networks Using Multi-User Co-Operative Motion Mechanism

K. Pazhanisamy[1,*] and Latha Parthiban[2]

[1]Department of Computer Science,
Pondicherry University, Pondicherry, India
[2]Department of Computer Science,
Pondicherry University Community College, Pondicherry, India

Abstract

Routing protocols communicate route information as well as information about broken routes. It provides three primary benefits: route failure avoidance, load distribution, and load correction. The main aim of this research work is to design efficient algorithms for secure route creation and route failure recovery thereby improving the effectiveness by reducing communication overhead to give better QoS in mobile ad hoc networks. Secure route creation algorithm using pseudo-random number generation (PRN) to provide an efficient secure route creation is proposed to restrict the insider or outsider attack in MANETs. Data packets may be lost while using time-dependent data delivery (which does not route them to specified destinations). Through routing, data packets that are not delivered will harm the overall amount of data sent. Due to data packet losses and route discovery failures, there are additional problems. The route identification process must be moved away from the originating node as a result of this requirement and an

* Corresponding Author's Email: kpsamy09@gmail.com.

In: Frontiers in Quantum Computing: New Research
Editor: R. Anandan
ISBN: 978-1-68507-816-4
© 2022 Nova Science Publishers, Inc.

effective route identification method has been proposed. A novel multi-user efficiency realization method using polar based coordination method for improving system efficiency is proposed. A cooperative motion ant colony optimization (CMACO) method is proposed to ii improve the suggested routing method to enable multimedia communications on mobile ad-hoc networks. In MANETs, each node connects using predetermined parameters with no central authority. This feature makes MANETs more vulnerable to network intruders. However, wireless data connections make MANETs more susceptible to attacks, allowing hackers to masquerade as the web and observe current data transmission. Data security is critical while transmitting data to a data device. Many new security-based protocols assist in minimizing these vulnerabilities that are available to all kinds of MANETs. An efficient intrusion detection partial swarm optimization (IDPSO) and support vector regression (SVR) is proposed that improve anomaly detection and give additional information about attacks. The experimental results justify the significance and necessity of the proposed work with required further enhanced investigation.

Keywords: MANET, route creation, AODV, PSO, randon number generation

1. Introduction

A MANET is a default self-organizing set-up without any supervisor. The establishment of two nodes is not inside the range. Other additional nodes are needed to help as in-between routers for the data communication between the two nodes. Mobile nodes walk-in destination entry and interconnect with the dynamic exchanging network. Thus, repeated alteration of the network is a hard-hitting challenge for numerous foremost problems like robustness in routing protocol and performance degradation. Routing protocols are mainly used for optimal packet route determination and transfer of the jammed data between the source and destination. Routing protocols aim to exchange route information and collect information regarding the non-operational routes.

The main beneficial features of the routing protocols are route failure, load balancing, and load repairing. Based on demand, the routing protocols won't exchange routing data from one point to another. Instead, they determine the route during the necessary condition for the data communication among two nodes [1]. Due to the dynamic change in the ADN setup, the links among nodes are never-destination. The time node can't direct packets to the particular destination node, and so the data packets possibly will get lost.

Missing data packets will affect the performance of routing in overall data communication. These data packet losses and route reply loss get additional problems. For this purpose, that source node requires re-transferring the route detection process.

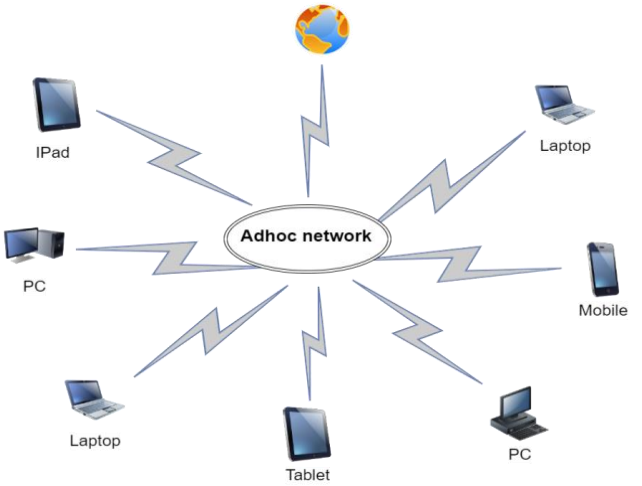


Figure 1. MOBILE Ad-hoc network structure.

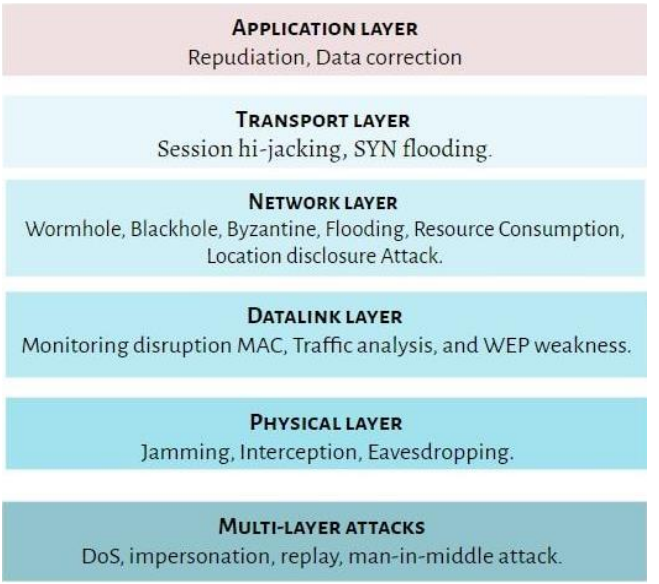


Figure 2. MANET layer attacks.

MANETs will be working without central supervision where the nodes interconnect with each other based on standard parameters. This specific feature makes MANETs more susceptible to being attacked by a hacker inside the network. Wireless data links create MANETs more vulnerable to attack, allowing the hacker to impersonate the web and acquire access to the current data communication. Furthermore, mobile nodes inside the wireless range link can eavesdrop and then simulate to the network [2, 3]. Therefore, a secure approach for broadcast and data transfer in MANETs is needed, a moderately challenging and essential concern due to increasing threats of data attack in Mobile Networks. The ongoing challenge is data security. It is critical to provide secure data transfer and communication to the mobile user. The system has to recognize different kinds of attacks and their unique effects on the MANETs. Attacks are categorized into two categories passive and dynamic attacks. Passive attack defines the data transmission eavesdropping, and it is difficult to identify. Next, a dynamic attack includes generating or altering a falsified data stream. All kinds of MANETs have no security and can be easily susceptible to every type of attack. Numerous novel security-based protocols were developed to remove these vulnerable attacks and acquired MANET's securities to handle those attacks. Figure 1 shows the mobile Ad hoc network structure.

Ad-hoc networks (ADN) are extensively settling in narrow areas like the hotel, household, cafeteria, and additional people gathering public places like business centers, workplaces, etc. for in-house and commercial applications. Over the widespread use of 6G, the rising demand for enormous transmission speed, little latent, and quality of service (QS) was the central alarm that desires extra investigation. To reassure high QS, efficiency is an essential factor to progress the broadcast speed, i.e., the data per unit time in the entire wireless communication system (WCS).

They have been accomplished through broadcast expansion bandwidth [4, 5] techniques. The above said approaches for every interval intake a significant amount of television data or more communication overhead to protocol standards. The mobile user (MU) co-operation information has been confirmed as a less cost protocol to deliver more efficiency, permitting dynamic users to co-operate by transmitting the resources from source to destination [6]. That mobile user whose QS is not guaranteed can ask the remaining users to perform as mediators to send. The user co-operative motion technique establishes the effectiveness in all the applications, which practices the co-users signal to travel to a good location for filtering the broadcast speed

[7, 8]. Later moving towards the most acceptable position, are minimized interruption.

The S/N ratio is the vital efficient progress factor. Unfortunately, the prior approaches are dependent on different user motion mechanisms; they are not usable in a motion space context. A significant additional benefit of the multi-user motion methodology is that it will resolve the NP-hard issue and a tremendous amount of communication overhead in previous approaches. Figure 2 shows the MANET layer attacks.

2. Literature Survey

By-pass routing is demonstrated as an on-demand limited retrieval procedure for MANET. The bypass AODV practices cross-layer MAC announcement to limit mobility correlated link failure and creates by-pass among the fractured link destination nodes through another node while keeping the remaining nodes of the route. The by-pass AODV is improved related to the old-style AODV; simultaneously, the error retrieval stage is removed, decreasing routing storage complexity and packet droplet ratio. The bypass AODV transfers the packets through the recently created by-pass route to avoid packet drop. Simultaneously, the system performance of by-pass AODV is top at raised node density once the distance jammed between the ending nodes is better than or equivalent to three stages. A small thickness of nodes where node inter-connectivity is short don, by-pass AODV is inappropriate due to an accident incident. Movement forecast and direction-finding are applied to overcome route failures by attaining.

DSR procedure's [9] route detection/route response cycle. To ensure the route-based cache of a packet source node is correct, it needs to be validated. This node creates abuse packets. The source node may create route detection packets to every or all surrounding node. However, if no legal path is present in the route-based cache, a route demand packet is sent to the network. Due to the little data measure communications overhead associated with source routing, DSR is a single drawback. Therefore, each route-based cache gathers the addresses of each visited node; RREQ packets will turn out to be enormous as they broadcast from source to destination in the network. Routing information can even rise sufficiently to go beyond the corresponding data's efficiency. These problems limit the network's suitable thickness and consequently its scalability.

AODVR procedure [10] is a source node that needs to direct data to a destination point for which it does not have a route transmission of an RREQ packet through the network. Every individual node getting this packet renovates their data for the source node. Hence, not like DSR, this method will not use route-based caching. As an alternative, every node preserves simply the subsequent hop's address in a direction-finding table; in addition, these routing-based tables have restructured the manner completely along the RREQ broadcast path. So for RREQ, the BCP bottleneck approach used increased system performance and advanced data-based packet transmission speed.

The scientists don't think over the different channel conditions while relating the BCP blockage process. Most of the techniques did not worry about other channel permission circumstances and substantial bandwidth or administration properties to improve efficiency. The cooperative, mobile user motion can reduce the data source consumption for the service provider to raise the QS. [11] developed a resourceful data sending procedure.

2.1. Research

2.1.1. Gap Identified from the Literature Survey

Insecure route creation techniques, most of the researchers do research is focusing on high-destination secure routing. But specific gaps are found like

1. Several hops cause performance losses because of the high error rate, network congestion, and potential association failure.
2. Routing and increasing the link failure is the disadvantage in ADN.
3. Local repair makes progress the route when a node distinguishes the destroyed link in a dynamic route by redirecting completely, and the proposed procedure consumes moderately more computation time.
4. Many ineffective routes are related to the route detection process, which tends to have a huge computation overhead.
5. DSR's main drawback is that the vast information measure overhead in-built in source routing.
6. AODVR's main drawback is the vast storage space and preserves the multiple addresses in the direction-finding table.
7. A multi-user motion system will tend to the NP-hard problem and huge communication overhead.
8. To help reduce the intrusion, the user must not be moving.
9. Service irregularity during the routing.

10. To minimize latency and to distribute bandwidth most efficiently, use least-possible latency and minimal bandwidth routing.

2.1.2. Objectives

This research work explains the following five aspects.

- Every node is equipped with a separate antenna and can't transfer packets simultaneously.
- Each node will have the same broadcast control. It requires minor alterations which yield data based on traffic.
- The data packets follow the multiple hop sequence. The topology will not get change with the node motion as they maintain strong relationships among nodes. The mediators assure that the transmission path remains unchanged.
- The Route Failure algorithm (RFA) for ADN development begins retrieval from route failure suddenly when route repair. RFA identified past route failure, and the deepest node searches for an alternative path around the defective area by selecting the main RREP that is the safe route and creates a new route to the perfect destination for transferring the data which are denied for some time delay.
- RSA developed a secure route construction system for Pseudo-Random Integer Generation (PNR). It used a random bit sequence integer function [12, 13]. Every node will produce a random bit sequence integer throughout the signaling period and inverse the random bit sequence integers order when reaching the destination node. The importance of this security system is to safeguard from any attacker or stranger node attacking the node of the wireless system. Furthermore, the random bit sequence integer function creates the web with extra safe and sound in place of MANETs with Quality of Service (QoS) and Intrusion Detection.

2.1.3. Contribution

This research was done in five phases:

Phase 1

RSA Algorithm developed a secure route creation system for Pseudo-Random Integer Generation (PNR). An RSA used a random bit sequence integer function. Therefore, every node will produce a random bit sequence integer during the signaling period and inverse the random bit sequence integers order

once getting the destination node. The importance of this security system is to safeguard from any attacker or stranger node attacking the node of the wireless system. In addition, the random bit sequence integer function creates the web with extra safe and sound in place of MANETs.

Phase 2

At the route repair, the Route Failure algorithm (RFA) for ADN introduces retrieval from route failure. Route Failure algorithm (RFA) organizes in every node gathers RREP safe in the high-destination security, it becomes activated throughout route failure. An innermost node will choose the primary RREP, which is the safe route in the RBT, and then start a unique route to the target.

Phase 3

The earlier method uses a single motion user and doesn't have multi-user co-operative motion arrangements. Static users developed an Efficient Multi-user Efficiency realization technique (EME) with the help of geographical location data. Simulation analysis results show that the efficiency of EME is improved when related to other approaches. Communication time of the EME technique is few when associated with the other methods.

Phase 4

QoS provisioning in such networks is complicated by issues like limited access to resources and the use of insecure media. This technique was created by Ant Intelligence to find the most efficient path for data transfer from origin to the destination while complying with the QoS constraint. The main objective of QoS is to choose the most optimal route to satisfy QoS resource requirements while also maximizing the usage of available resources.

Phase 5

It is difficult to detect new attacks when there are no known signatures, which is a major drawback of the system. So how IDPSO and SVR may enhance malware detection techniques and provide more information about attack sources and types. Here, how signature attacks are made on a simple algorithm to identify numerous well-known attacks once an anomaly is found.

3. Proposed Algorithm

Every node is allocated an exclusive ID before building a wireless system for data transfer among nodes. Once a wireless web is created, node (S) desires to transfer by node (D) in the network topology. The node (S) performs as a transfer node. It will transmit data throughout the signaling period by unique node ID and produce a random integer, e.g., when the node transferred this data communication to its neighbors. Receiving node for (E) initially validates that it has an ID of sharing node in its routing table. Figure 3 shows the secure route creation.

If receiving node (E) is not the destination point, it transmits the data once more. The node (E) will communicate data with its unique ID, and it recognizes that node (S) generated a random integer (a_1); thus, once computation through a random integer, it will produce an integer (a_2) as soon as the destination node (D) will get the unique ID with a random bit sequence (a_1,a_2,a_3…a_k). When D returns, all the random bit sequence integer generating orders will be "inverted" and repulsed.

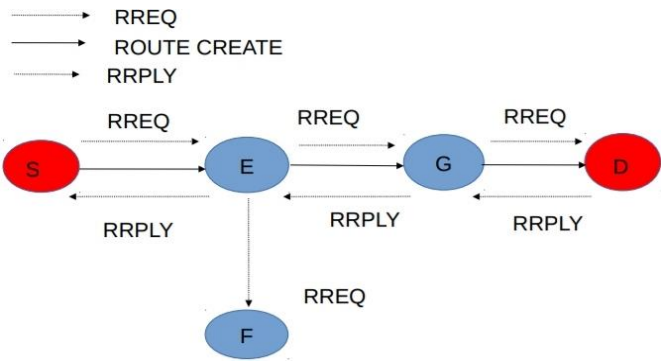


Figure 3. Secure route creation.

The unique ID and a random bit sequence integer are provided by the destination point node (D). node (G) sees that it generated a random bit sequence integer a3 then agrees to it. A similar process is monitored in the entire network till it touches the source node (S). Next to this signaling procedure route is set between the source node to the network node in a protected location. Thus, data is communicated safely without any attack.

4. Proposed Work

4.1. Secure Route Creation for ADN Using RFR Algorithm

This concept is often referenced when new strategies (Figure 4) come. After node (S) establishes a wireless network connection, it must then connect to another node (D) inside the setup. An instance of the transmitting node known as the S connects and sends a message across the network, displaying an ID value unique to the nodes. An S randomly generates a range of bits to be an example for its neighbors when it broadcasts. In addition to ensuring an active (E) 1st route, a receiving node has an associated ID for the destination server in its traffic database.

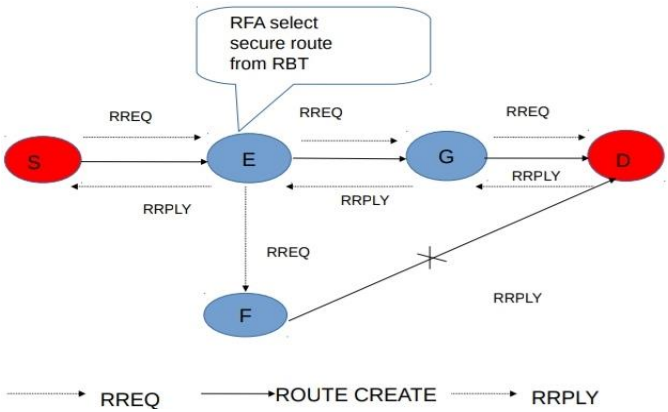


Figure 4. Secure route creation for ADN Using RFR algorithm architecture.

If the route to the destination is already present in the routing table, it will accept the response message sent by the originating node. The first time it broadcasts the message, the message discards. The node (E) may send a message using its unique ID, and it knows about the node (S), which produced a random bit range (a1). Once the destinations node D has received the random bit sequence (a1,a2,a3, etc.) with a unique ID, the trip may begin. Any random bit series range generating order may be sent back and then reversed to D. Once node D (the destination) provides its unique ID, create an entirely random bit sequence for D and send it. After having completed the random bit series range A3, node G will accept it. The whole network follows the same process, except the supply node (S). After the secure communication channel setting between the supply nodes and the network node, no unauthorized

person or assault is allowed throughout this algorithm. While system RFA is implemented in each node, RBT responds with RREP packets in the most secure route from relevant downstream nodes. If there is a route failure, the first RREP in the route block is selected. The RREP continues running until the end of the route block. With the supply node modified, the routing table of all relevant nodes is updated. The Network's Route Failure Recovery algorithm is enforced and assessed (NS two, version 2.32). Route may significantly improve the simulation of wireless networks with the usage of the NS2. Price NS2 may be an effective alternative to network styles already in use.

4.1.1. RFA

Design

RFA performs under similar mobility and traffic situations to determine their performance. AND is a simple form that does not contain any of the other search improvements, including the expanded rings searching and the localized route repair. By using this, an equal comparison of outcomes is done. This is the percentage of data packets produced by CBR sources that get transferred to their destinations. This measures the protocol's route discovery capabilities. Figure 5 shows the design of RFA in ADN. Because of its overhead, a significant routing burden is placed on each message. Every station along the route tallied each broadcast. In reality, it is the inclusion of certificates and signatures in the packets that causes the problems.

Algorithm

- Route failure detected
  Else
- Data packet transmitted
- If route failure is detected, RFA is activated

This metric may interpret in two ways: it either counts control packets and the overheads (if any) relative to data packets or measures control packets while accounting for overhead. When a trusted certificate has been revoked, the trusted certificate's revocation certificate may sometimes be the only link between two different sections of the ad-hoc network. The server only needs to use one connection to access both partitions of the network. Because the

certificate has been revoked, the two sections of the network cannot communicate with each other. Table 5 shows the simulation setup of routers.

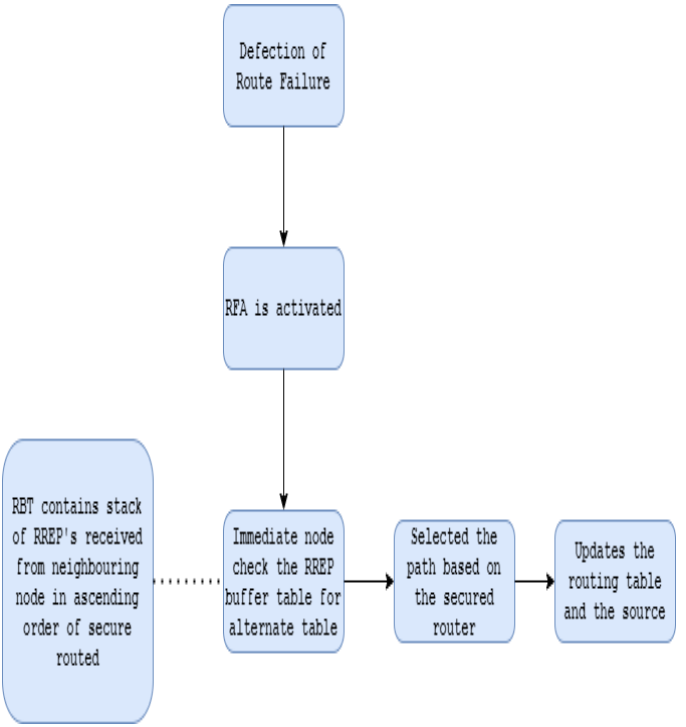


Figure 5. Secure route creation for ADN using RFR algorithm design.

Table 1. Secure route creation for ADN using RFR algorithm simulation setup

Parameter	Value
Test Area	WLAN
Number of Nodes	45
Transmission Bandwidth	2.0 Mbps
MAC	802_11
Traffic Type	CBR
Packet Size	1500
Initial Energy	50

Figure 5 shows the secure route creation for AND using the RFR algorithm.

4.1.2. Performance Analysis

RFA's throughput is compared to the basic AODV routing protocol concerning average packet ratio value. Packet delivery simulation findings show that when AODV is used in combination with RFA (Figure 5), the results will be over-stated when AODV fails due to a network fault. As a result, the RFA shows better results during connection failure than AODV [14, 15]. As more failed nodes are present, the routing protocol using RFA tends to be better than AODV.

Usually, knowledge packets have to send delayed because of the time it takes to traverse the path to the destination. According to Figures 6 and 7, AODV with RFA produces much greater outputs than those produced by AODV when a link failure occurs.

Participation without Authorization

Participation in the ADOV network is limited to approved keystrokes, as sent by the entity responsible for issuing such keys. In reality, single-hop VPN installations are in widespread use on university campuses, such as at the University of Massachusetts. Authentication mechanisms for using a trustworthy certificate authority have many means. gives a substantial list of options. However, there are situations when using several redundant authorities is advantageous.

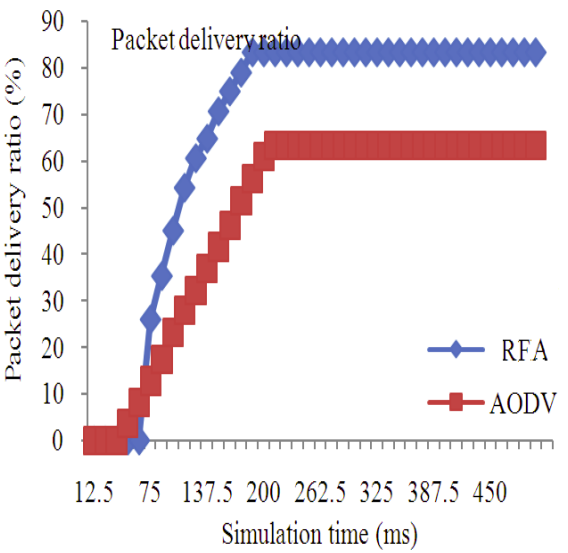


Figure 6. Packet delivery ratio.

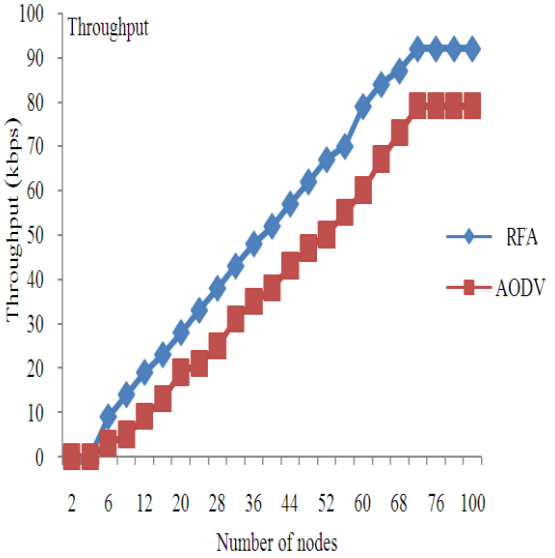


Figure 7. Throughput.

Route Signaling Spoof

Nodes cannot spoof other networks in route creation, while only the source node may use its private key to sign the route. When responding to route discovery, the source node includes its certificate and signature, allowing it to be the only one who may answer. With this in place, the attack known as an impersonation attack is avoided.

Routing Message Alteration

Any field of an RDP and a REP packet is left untouched in the routing process. To identify and reject a changed packet, intermediate nodes along the route must keep track of all of the packets in transit. The node that modified a packet may send a new version of the packet, and if the modification is discovered, the altered packet will be rejected. The fear of changing packets having damaging effects on other nodes in the network is not addressed.

RFA

RFA performance is related to AODV routing-based protocol on behalf of its packet delivery quantifiable relation, throughput. The simulation effects of packet delivery quantifiable consideration of AODV through RFA routing procedure more efficient when compared to older AODV routing procedure

through link failures. It's moreover resolute that the RFA is moderately reliable all through the link failures when compared to AODV. Once there are many failure nodes, the routing procedure through RFA leads to a better RFA related to the AODV. The RFA attains higher output when put next to the opposite case since the alternative path selected by the RFA is resulting in higher output.

4.2. Multi-User Efficiency Mechanism

4.2.1. Efficient Cooperation Model

Broadcasting information-sharing system is based on efficient responsiveness. When it comes to BCP downlink, the ACO technique is utilized to achieve the organizing and cooperative connection adaption. Nevertheless, the broadcasting frequency does not support active channel organization. A new method is created to allow many users to converse simultaneously and provide appropriate broadcast parameters for their requirements. Broadcast is a wide-ranging system that works in tandem with many different technologies to provide efficient user system performance. Noise is not measured, nor is the channel balance and bonding. They are incapable of being used in real-world applications. Node completed the BCP congestion process (increasing the system quality and enhancing the packet data broadcast speed) with the block size response of frames accumulating and modification. Deductive reasoning is not used while using the BCP bottleneck method. To use all the available techniques, regardless of the access scenario, and to make the most efficient use of bandwidth or other assets, a novel method is proposed to get maximum resources.

4.2.2. Multi-User Co-Operative

Method

A multi-user cooperative motion planning system creates a collaborative environment for many users as shown. ADN modeling using the acronym AH, while S is just S. In other words, S plus D would be the collection of dynamic nodes that make up D. In addition to |S| nodes, there are |D|, equal numbers of dynamic and static nodes. A node is connected to every other 57 nodes in the whole network, and every node connected to the overall system is assigned a topographical location. Figure 8 shows the Cooperative multi-user motion

However, dynamic nodes connected to the system may shift positions within a partial 2-dimensional circle. This circle has a radius of Rc.

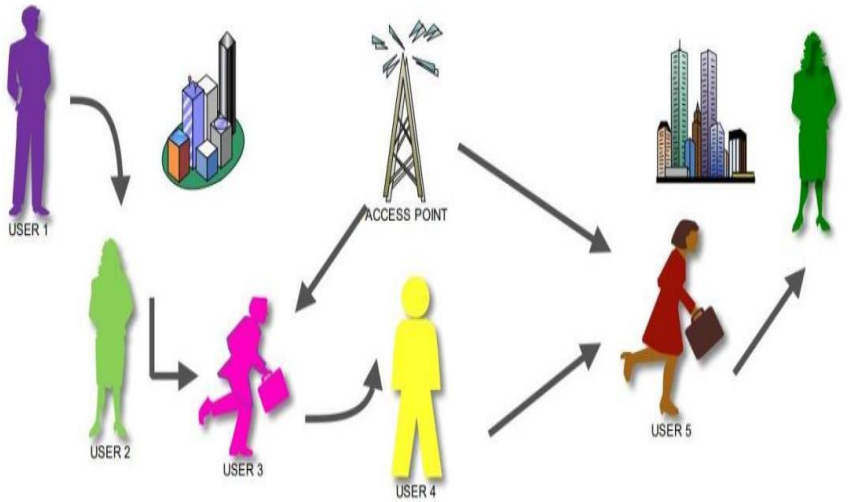


Figure 8. Cooperative multi-user motion.

Figure 8 illustrates the multi-user cooperative mobility system, which has users from several transceiver sets exchanging data through mutual usage. Accordingly, the following assumptions are made:

- For every nodule, there is an antenna, and it cannot transmit packets at the same time.
- A nodule is also the case due to the almost equal antenna efficiency. It features tiny radius merger adjustment; along with this, the data-based traffic is generated.
- It happens on a single hop, and each packet takes two hops. If there are stable connections between nodules, then nodule mobility won't alter the flow of information along the broadcast route.

Motion with Broadcast

Figure 9 illustrates the collaborative motion communication application configuration. First and foremost, the static and dynamic nodules distribute across a limited spherical area at random. Additionally, nodules exist that are far from ACP. 59 Because they are requesting new active collaborators to be intermediaries and relay the information-based traffic, they will also need

participants to function as mediators. The idea was that the virtualized necessary controllers distributed across the ACP, such as google maps-based navigation, would execute the suggested system and construct the positions of each dynamic nodule. To display each dynamic nodule's geographical position, the ACP routes communications from the controller to the PAML. Frames include controller information as the header for data communication traffic, which helps ensure a more accurate controller lookup when multiple hops or dual-hop nodules are present. This control communications method is similar to 'beacon.' Using the user datagram procedure, the mediators send communications during the broadcast period. To generate an uplink-based calculation, first, get the data from all users' activity that causes uplink data (ACP); this is the data source used for all connections. The ACP (typical vocal) broadcasts the total amount of data received every time it is active. This technique allocates each nodule a place in the restricted-space sphere in the most efficient way. Improving the scheme's effectiveness by getting to the site with the best terms is efficient. The high SNR is critical for topographical areas that remain highly efficient.

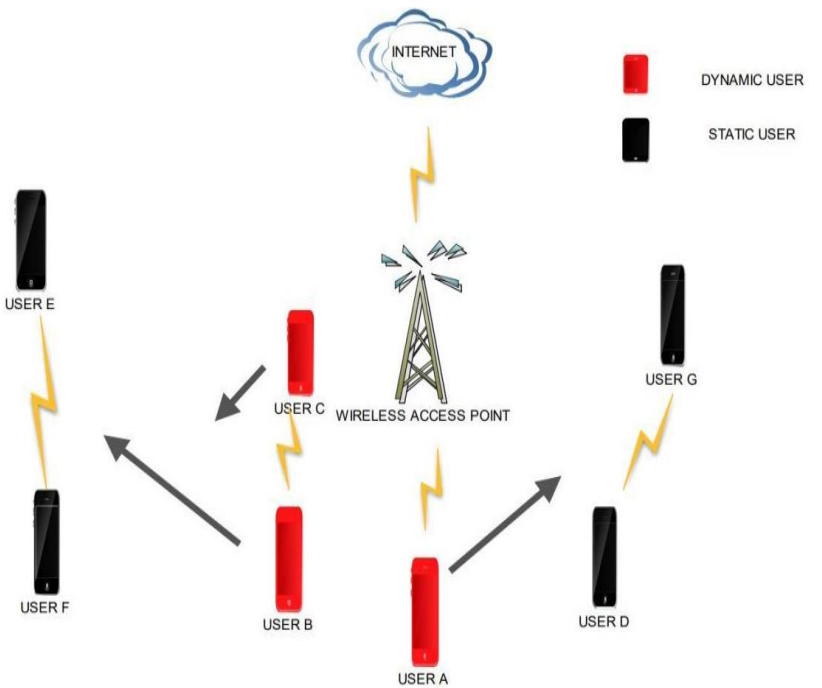


Figure 9. Broadcast method for multi-user co-operative motion.

The system's application configuration for cooperative motion data transmission, static, and dynamic nodes are subjectively positioned in a limited sphere-shaped space. Many nodes are staying away from ACP; also, they demand extra dynamic agents to turn out to be mediators and interconnect the data-based traffic. The ACP leads the data control data communications by the controller to display every single dynamic node's topographical location.

4.2.3 Working Mechanism of Cooperative Method

To define the positions and also the distances between twofold nodes, use polar coordinates. Let the location of the nodule $h_i \in AH$ here describes as h_i (ρ_i, θ_i) having a polar diameter of ρ_i, where ρ_i indicates the opposite dimension, $\rho_i \leq Rc$, and θ_i the polar angle, represented by I, is equal to the position angle of the nodule. AH position pair nodules as AH(ρ, θ), AH(ρ, θ) = $\{h_i (\rho_i, \theta_i) | h_i \in AH, \rho_i \in \rho, \theta_i \in \theta\}$, $|p| \leq Rc$

$$\text{Static - dynamic nodules } s_i \in S,\ d_i \in D,$$
$$\text{Sets of position are} S\ (\rho_S, \theta_S) = \{S_i(\rho i, \theta i) \vee S_i \in \rho_f, S_i \in \theta_f\}$$
$$D\ (\rho_M, \theta_M) = \{M\ (\rho_i, \theta_i) |\ M_i \in \rho_M, M_i \in \theta_M\}$$

Dual arbitrary nodes of $h_1 (\rho_1, \theta_1)$ and $h_2 (\rho_2, \theta_2)$,
Then $d_{h1,h2}$ is defined by

$$\begin{aligned} d_{h1,h2} &= \|h_1 (\rho_1, \theta_1) - h_2 (\rho_2, \theta_2)\| \\ &= \{ |\rho_1 - \rho_2| ,\ \theta_1 - \theta_2, \theta_1 \neq 0,\ \theta_2 \neq 0 \\ &= \sqrt{(\rho_{12} + \rho_{22}) - 2\rho_1\rho\ 2\cos(\theta_1 - \theta_2)} ,\ \theta_1 \neq \theta_2 \end{aligned}$$

Theorem 1

$\lambda_{vr}(\eta_i)$ falls in the range of the C:r (θ) = Rc, η_i is inside the C, and $\eta = \gamma(\eta_i)$ represents the PPT distribution.

$$[\gamma(\eta_1) = k] = [\lambda_{vr}(\eta_i)]^k exp\left(\frac{(\lambda_{vr}(\eta_i))}{m!}\right)$$

Alone, isolated locations such as $\gamma(\eta_1)$, $\gamma(\eta_2)$……$\gamma(\eta_s)$ have no interconnecting factors. Thus, the static nodules inside a particular location are entirely random, with a specified probability of points $\gamma(\eta_1)$ per unit area, which means that their presence does not change whether the issues generates

within a region, a location, or an outer circle. No arbitrary variable exists between two static nodules of the space; it is governed only by the parameters and functions that have been randomly scattered throughout the space, supporting an exponential-based distribution mechanism called γ. Because there are two static nodules in the restricted space, there is a chance of a η_i.

$$p[\gamma(\eta_i) = 0] = e^{(\lambda_{vr}(\eta_i))}$$

It calculates as follows: The cumulative distribution function S ($l_{si,Sj}$) obtaining as:

$$S\left((l_{si,sj}),\gamma\right) = S(l_{si,sj} \leq r)$$
$$1 - p[\gamma(\eta_i) = 0]$$
$$1 - e^{\lambda\pi r^2}$$

Given that nodes S_i are all stationary, $S(l_{si,sj} \leq r)$ determines the likelihood of separation amongst these pairs been of nodes. Thus, $S_i \leq r$ using the properties of here, to estimate $E(l_{si,sj})$ the expectation of the returns to several equity indices,

$$E(l_{si,sj}) = \int_0^{Rc} 2\Pi r^2 exp(-\lambda\Pi r)^2 dr$$
$$erS_i(R_c\sqrt{\lambda}\sqrt{\pi})_{2\sqrt{\lambda}} - R_c exp\left(-\lambda\pi {R_c}^2\right)$$
$$erS_i(\eta) = \frac{2}{\sqrt{\pi}}\int_0^{\eta} e^{t^2}\, dt$$

In addition to perusal PAML and MDAC frames, inter signal interference analysis for users who are not in PAML, including verifying the frame arrangement and sending packet conflicts in the link-layer MDAC frame representation. On the live television broadcast, data frame conflicts are evident. To hypothesize that data stream saturation has occurred. However, the issues with data frame conflicts don't affect a part of results, and that's why frames put restrictions on data usage in place.

From Figure 10: the efficiency comparison of the various protocol is obtained.

In Figure 11 the communication overhead of protocols is shown.

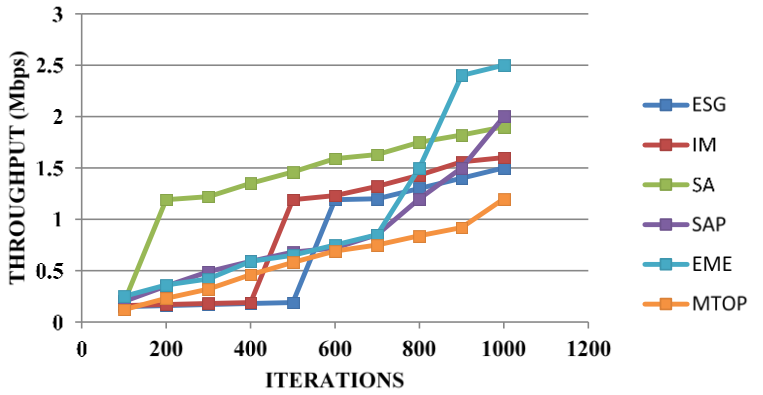


Figure 10. Protocols comparison.

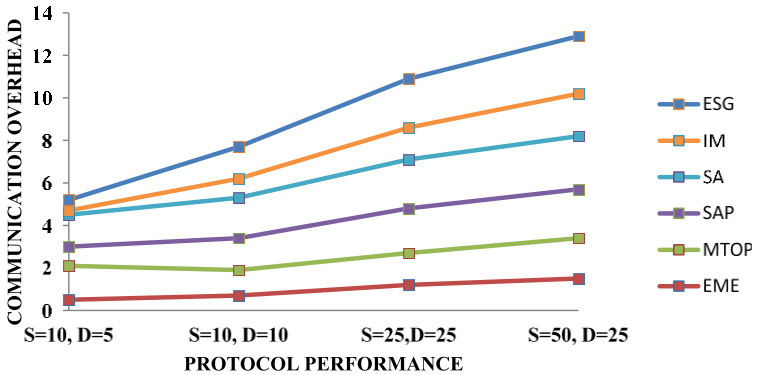


Figure 11. Communication overhead.

4.3. An Efficient Cooperative Motion - Quality of Service in Mobile Adhoc Networks

An Ad-hoc mobile network is a wireless network that may be assembled on-demand. Network status information is typically inexact in Ad-hoc networks, while in many other networks, service quality is far more difficult to guarantee. Routing is the act of moving information from one location to another by way of a network. The complexity of a MANET increases because of many characteristics, such as mobility, time-varying QoS, and limited energy and resources. It is only possible to find the route if the node cooperates. Adding constraints like limited resources, untrustworthy media,

complicate QoS provisioning in these networks. A route to the data delivery path from origin to destination has been created to provide data QoS control. The main aim of QoS is to choose an optimal route while ensuring the effective use of resources. In the MANET routing protocol, no consideration is given to power-conscious routing or QoS delivery. CMACO and Ant Colony algorithms have proven effective in the construction of MANET routing algorithms. This paper presents new QoS algorithms for the MANET. CMACO and AODV combined identify different stable routes between source and destination node.

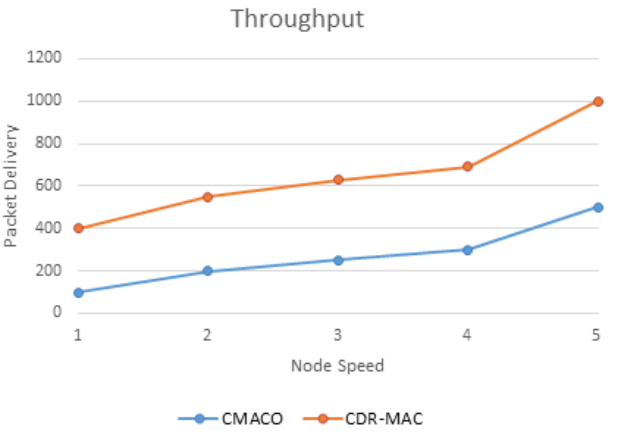


Figure 12. Throughput of CMACO and CDR-MAC.

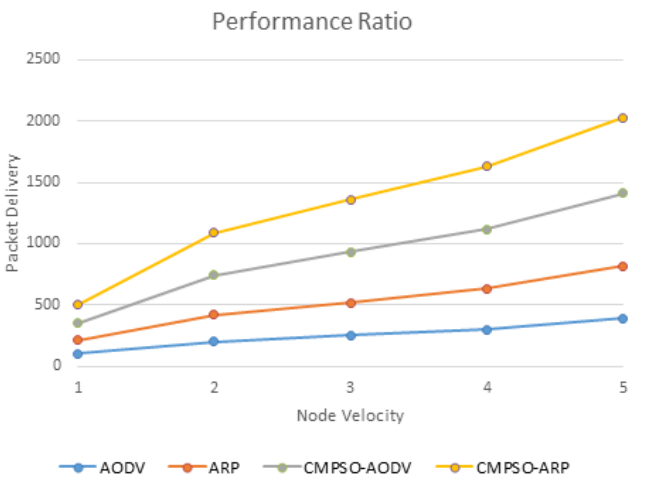


Figure 13. Performance ratio.

Figures 12 and 13 demonstrate simulation setup throughput and packet delivery, and this highlights the ideal option (or almost optimal solution). Though the delay curve is greater, the delay is less than the genetic approach. CMACO varies little from the dynamic programming in the postponed fluctuation.

4.4. Mobile Ad-Hoc Network Intrusion Detection in Cooperative Motion

The increased use of wireless devices has sparked the emergence of mobile ad hoc networking (MANET). MANETs were vulnerable due to their open, constantly changing network architecture, sharing methods, absence of a protection measure, and a cohesive layer of attack. However, a signature attack on daily operations often produces fake warnings. Another drawback is the inability to detect new attacks without already-existing signatures. We discuss our attempts to build a capability for MANET intrusion detection (ID). In light of our prior research on outlier identification, we see how partial swarm optimization and support vector regression may provide additional information regarding attack sources and types. SVM and PSO may use a general formula to identify various well-known attacks when anomalies are observed.

SVMs (Support Vector Machines) were created first to classify issues. SVR is an SV model. A non-linear mapping function for training data was one of the main objectives of the SVR model. the input location, the actual output, and the quantity of the data show the training data set

$$a = f(b_i) = \omega^M \varphi(b_i) + z$$
$$f(b) = \sum_{i=1}^{n} (\alpha_i - \alpha_i^*) m(b_i, b) + z$$

$\alpha_i \wedge \alpha_i^*$ consists of the Lagrange multipliers and are a kernel function. When applied to the SVR input space, the kernel function produces a nonlinear decision hypersurface. The Stochastic radial basis (RBF) kernel, the most commonly used kernel, conducts a mapping function between the spatial domain and a high-dimensional space. Still, it is also simple to construct, making it ideal for addressing nonlinear issues; it was the motivation for using the Stochastic RBF kernel in this research:

$$k(b_i, b) = exp(-\sigma \vee b - b_i \vee^2)$$

Figure 14 and 15 shows the occurrence and comparison chart of the proposed work. This study makes use of the raw data from the 2019 intrusion detection competition. This database includes many military-militarized intrusions, and it serves as a baseline for intrusion detection in general. Probes and DDoS attacks make for the bulk of all attacks. While compromises are exciting and risky, they are grossly underrepresented. There are 42 properties in all, including one class label for each connection record. We regard all 24 distinct attack types as one unique assault group. It required to process a data collection of size N. Numerical values were changed to nominal values, although those values were all discrete (integers). N-rowed and M-columnar representation of vector X (attributes). There are eight discrete values and thirty-three continuous values in the database. Table 1 outlines the KDD intrusion detection design.

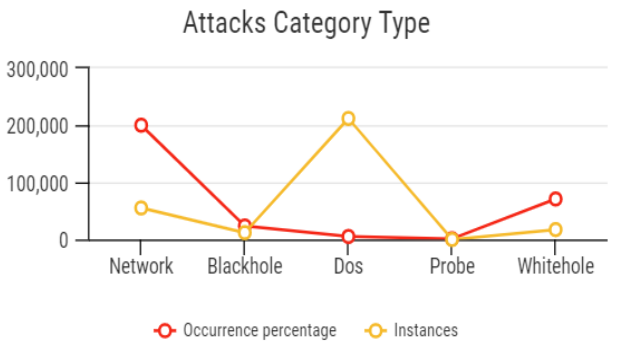


Figure 14. Occurrence of attacks percentage.

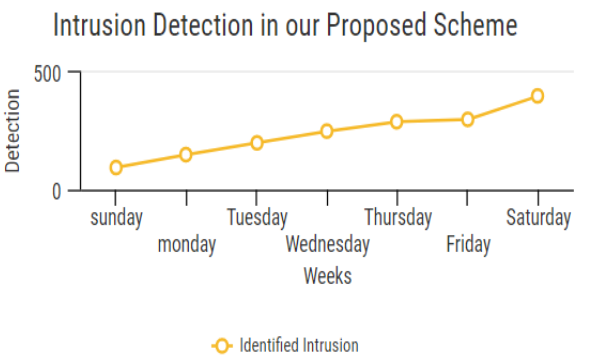


Figure 15. Proposed scheme detection.

Conclusion

ADN to acquire over link failure termed the Route Failure algorithm (RFA) with AODV routing protocol. The RFA algorithm's performance merged with the AODV routing protocol is related to the earliest AODV in terms of packet delivery relation, production, and establishing higher efficiency in every specific aspect. This may be achieved because the RFA is triggered improvised through link failure, thus decreasing knowledge-based packet failure.

The co-operative multi-user motion technique depicts the categories of static node placements. Then, ADN calculated the time required for the system's whole operation. Finally, ADN determined the calculation time needed. In the future, the proposed research is ext-destination to design a novel secure route-finding protocol that uses minimum security factors to avoid storage overload in the network.

References

[1] Singh, T.P., S. Dua and V. Das Energy-efficient Taneja, S. and A. Kush, 2010. A survey of routing protocols in mobile ADN. *Int. J. Innov. Manage. Technol.*, 1: 279-285, 2012.

[2] N. D. Han, Y. Chung, and M. Jo, "Green data centers for cloud-assisted mobile ADNs in 5G," *IEEE Netw*., vol. 29, no. 2, pp. 70–76, Mar. 2015.

[3] E. Huang, W. Hu, J. Crowcroft, and I. Wassell, "Towards commercial mobile ADN applications: A radio dispatch system," in *Proc. 6th ACM Int. Symp. Mobile Ad Hoc Netw. Comput*. (*MobiHoc*), 2005, pp. 355–365.

[4] S. Moh and C. Yu, "A co-operative diversity-based robust MDAC protocol in wireless ADNs," IEEE Trans. *Parallel Distrib. Syst*., vol. 22, no. 3, pp. 353–363, Mar. 2011.

[5] Taneja, S., A. Kush, and A. Makkar, Experimental analysis of DSR, AODV using speed and pause time. *Int. J. Innov. Manage*., 1: 453- 458, 2010

[6] Q. Zhang, Q. Chen, F. Yang, X. Shen, and Z. Niu, "Co-operative and opportunistic broadcast for wireless ADNs," *IEEE Netw*., vol. 21, no. 1, pp. 14–20, Jan. 2007.

[7] F. Chen, H. Zhai, and Y. Fang, "Available bandwidth in multirate and multihop wireless ADNs," *IEEE J. Sel. Areas Commun*., vol. 28, no. 3, pp. 299–307, Apr. 2010.

[8] Deepa, N & Pandiaraja, P 2020, "Electronic healthcare system data privacy-preserving efficient file retrieval from the cloud service provider using attribute-based file encryption", in *Journal Ambient Intelligence and Humanized Computing*, DOI: 10.1007/s12652-020- 01911-5, Annexure I, Impact Factor: 4.594.

[9] F. Wang, L. Ruan, and M. Z. Win, "Co-operative network operation design for motion-aware cloud radio access network," *IEEE Trans. Wireless Commun.*, vol. 17, no. 12, pp. 7819–7833, Dec. 2018.

[10] Argyriou, "Cross-layer and co-operative opportunistic network coding in wireless ADNs," *IEEE Trans. Veh. Technol*., vol. 59, no. 2, pp. 803–812, Feb. 2010.

[11] Z. Hu, X. Wen, Z. Li, Z. Lu, and W. Jing, "Modeling the TXOP sharing mechanism of IEEE 802.11ac enhanced distributed channel access in nonsaturated conditions," *IEEE Commun. Lett*., vol. 19, no. 9, pp. 1576–1579, Sep. 2015.

[12] B. Baron, P. Spathis, M. Dias de Amorim, Y. Viniotis, and M. H. Ammar, "Motion as an alternative communication channel: A survey," *IEEE Commun. Surveys Tuts*., vol. 21, no. 1, pp. 289–314, 1st Quart., 2019.

[13] M. Aajami and J.-B. Suk, "Optimal TXOP sharing in IEEE 802.11ac," *IEEE Commun. Lett.*, vol. 19, no. 7, pp. 1141–1144, Jul. 2015.

[14] Chaintreau, P. Hui, J. Crowcroft, C. Diot, R. Gass, and J. Scott, "Impact of human motion on opportunistic forwarding algorithms," *IEEE Trans. Mobile Comput*., vol. 6, no. 6, pp. 606–620, Jun. 2007.

[15] Cigdem, S. and R. Kravets, 2006. By-pass routing: An on-demand local recovery protocol for ADN. *Ad Hoc Netw*., 4: 380-397. DOI: 10.1016/j. .2004.10.004.

Chapter 10

Pattern Recognition Accuracy of Echocardiogram Images Using Deep Learning Techniques

Ilavazhagi Bala[1,*] and S. Latha Parthiban[2,†]

[1]Department of Computer Science, Bharathiar University, Coimbatore, Tamilnadu, India
[2]Department of Computer Science, Pondicherry University Community College, Pondicherry, India

Abstract

Echocardiogram is a highly designed system employed to discover heart-related diseases. An echocardiogram exploits an ultrasound to portraits the occurrence of diseases. To suggest the perfect remedy for the patient, precise recognition is the main task in clinical diagnosis and decision making. But, the pattern detection from the echocardiogram images is a difficult task with the presence of noise. The noisy pixels use more space introducing the computation complexity. This leads to affect the image quality and accuracy of detection. Several methods were employed to solve these problems, but, the error and time complexity were not minimized. To handle these difficulties, two new techniques are designed to improve pattern recognition. In addition, noisy pixel eradication is also accomplished in the proposed research work to lessen the complexity. Frost Filterative Fuzzified Gravitational Search based Shift Invariant Deep Structure Feature Learning (FFFGS-SIDSFL) technique is developed to enhance pattern recognition. The FFFGS-SIDSFL technique includes input, hidden units, and output units for accurate

* Corresponding Author's Email: ilavazhagibala@gmail.com.
† Corresponding Author's Email: lathaparthiban@yahoo.com.

In: Frontiers in Quantum Computing: New Research
Editor: R. Anandan
ISBN: 978-1-68507-816-4
© 2022 Nova Science Publishers, Inc.

pattern recognition. Echocardiogram videos are taken as input and it is divided into frames in the first hidden layer. Preprocessing, feature extraction, and pattern recognition are employed in FFFGS-SIDSFL to attain the objective of pattern detection. To start with FFFGS-SIDSFL, speckle noise eradication is accomplished in hidden layer two by enhanced frost filtering. During this process, the noisy pixels are replaced through the neighboring pixels. From that, the quality of the image is maximized. After that, an optimal mixture of features is chosen by the Stochastic Gradient Learning Fuzzified Gravitational Search algorithm at hidden layer three. The algorithm is improved with the aid of a fuzzy triangular membership function. With this, the features are selected and then extracted. By using extracted features, a feature vector is created. The output of hidden layer three is given to the output layer. Here, pattern recognition is performed using the Gaussian activation function. In FFFGS-SIDSFL, matching learned feature vector with training feature vector to detect patterns. This reduced the complexity with better accuracy.

Keywords: gravitational search, deep structure feature learning, pattern recognition, pattern recognition, ultrasound

1. Introduction

An echocardiogram exploits an ultrasound to portraits heart-related illnesses. The echoes are transferred into the images or video which is observed on the monitor. Due to the occurrence of noise, the processing of echocardiogram images and accurate pattern recognition is a complex one. In clinical diagnosis and decision making, precision recognition is a major step for providing the perfect remedy for a diseased person. By applying pattern recognition, heart disease is discovered with maximum accuracy. Recently, many works have been carried out for disease diagnosis using echocardiogram images. But, the noise removal was not effectively performed. Also, the pattern recognition with less time and high accuracy was failed to be attained. As a result, novel techniques are needed to be designed for obtaining enhanced pattern recognition accuracy with less time and noise.

1.1. The Objective of the Research Work

The main objectives of the proposed techniques are as follows:

- To improve the pattern recognition accuracy of echocardiogram images with less computational time
- To minimize the computational complexity of the Gravitational Search Algorithm is designed in the HEG-GS method.
- To enhance the performance of pattern recognition with higher accuracy, Frost Filterative Fuzzified Gravitational Search based Shift Invariant Deep Structure Feature Learning (FFFGS-SIDSFL) technique is proposed.
- To improve the peak signal-to-noise ratio and reduce the error rate, the frost filtering technique is designed in FFFGS-SIDSFL.
- To minimize the computational time and complexity, the Stochastic Gradient Learning Fuzzified Gravitational Search algorithm is developed to choose the optimal features.

1.2. Problem Definition

In medical diagnosis, pattern recognition provides efficient results for identifying the presence and absence of diseases. Several works have been introduced for medical image disease diagnosis using an echocardiogram. However, the computation complexity involved in pattern recognition was not reduced. Due to the noise presented in the images, the identification of actual borders of images is the major concern for attaining better recognition accuracy. Also, numerous machine learning techniques and optimization techniques are developed to perform pattern recognition. But, the error rate involved during the recognition of patterns was not reduced. Thus, it failed to increase the pattern recognition of echocardiogram images in terms of time and accuracy. The above issues are resolved by using the proposed methods.

1.3. Contribution of Research Work

The contribution of the research is described as follows:

A new Hierarchical Elitism Gene Gravitational Search (HEG-GS) method is developed to enhance the pattern recognition from echocardiogram images using optimizing the modular neural networks (MNN). The main objective of HEG-GS is achieved by using two various processes namely preprocessing and optimization. The echocardiogram videos are divided into several frames.

Here, the fore and fitness of the agent are measured. To increase the recognition accuracy, best fit and worst fit are computed. Then the velocity and position are measured for each agent. Finally, validation is performed through the average elitism and the process is repeated for all input images. This helped in achieving maximum recognition accuracy with minimal time.

Frost Filterative Fuzzified Gravitational Search based Shift Invariant Deep Structure Feature Learning (FFFGS-SIDSFL) technique is developed to improve the pattern recognition from echocardiogram images. FFFGS-SIDSFL has an input unit, three hidden units, and an output unit for deeply learning the feature vector to obtain accurate pattern recognition. The input echocardiogram videos are initially partitioned into frames. Enhanced frost filtering is employed in FFFGS-SIDSFL to remove the speckle noise in the image. The local deviation to the mean of the noisy image is measured to identify the noise pixels. Then the speckle noises in the electrocardiogram images are removed to get the noise-free images. Stochastic Gradient Learning Fuzzified Gravitational Search algorithm is applied to pick the optimal combination of features. The fuzzy triangular membership function is applied to enhance the Gravitational Search Algorithm. Stochastic gradient learning is used to obtain the best fit and worst fit from the population. With this, statistical features such as texture, shape, size, and intensity are extracted. The gaussian activation function is lastly applied in the output unit for matching the learned feature vector with the testing feature vector. Gaussian activation function provides binary outputs and thereby the diseased pattern and normal pattern are effectively-recognized. As a result, the FFFGS-SIDSFL technique gives accurate pattern recognition with minimal computational complexity.

2. Literature Review

A paired-views LV network (PV-LVNet) was designed in [1] using CNN and frame-level feature extraction. But, accurate clinical cardiac function estimation failed to achieve with lesser complexity in PV- LVN. A Neuro-Fuzzy based Genetic algorithm was introduced in [2] for the measurement of fetal growth by considering ultrasounds images. However, the designed optimization algorithm was failed to consider the computational cost in terms of time and complexity involved in pattern recognition.

The accurate categorization of pressure data was done to prevent heart disease. But, the performance of the disease diagnosis was not improved. An artificial neural network was presented in [6] to discover cardiovascular

diseases depending on the evaluation of echocardiogram parameters. However, the optimal feature selection and extraction were not focused to lessen the pattern recognition time.

An automated 3D echocardiography examination of Left Ventricular mass was studied in [7]. But, the noise removal and feature selection process was not carried out in the designed method. A convolutional neural network was designed in [8] to categorize the echocardiograms images taken from the videos. However, the filtering technique was not utilized to lessen the noise rate during the categorization.

A two-step transfer learning method was introduced in [9] through the filtering technique with more images. The method failed to analyze the performance of echocardiography translation with huge training images. Deep CNNs was developed in [10] using echocardiograms images to discover the local cardiac structures using obtaining local features. However, the designed CNN failed to present accurate results with enormous images.

2.1. Optimization Methods Based Disease Diagnosis

Optimization is the process of choosing the best element from the collection of available alternatives. The optimization comprised the quantitative disciplines from computer science and engineering. An optimization problem increased or reduced the real function by choosing the input values within the allowed set. The non-convex issues were addressed with the majority of available solvers for attaining better locally and globally better solutions. Global optimization is employed in mathematics analysis with a deterministic technique for guaranteeing the convergence of infinite time to achieve an optimal solution.

Multi-Detector Computed Tomography (MDCT) was constructed by Sudipta Ghosh et al., (2019). The health of one heart integrated doppler-based echocardiogram. The designed techniques needed expert supervision and costly clinical setups. An individual was employed to perform an efficient cardiac health diagnosis. A new impedance cardiography (ICG) technique was constructed for person cardio-vascular health identification. Recorded ICG signal assisted in feature extraction for cardiac health monitoring. But, the optimal feature selection and extraction were not concentrated to minimize the pattern recognition time.

The mid-level image representation was carried out by Otavio A. B. Penatti et al., (2015) for 2D echocardiogram ultrasound image plane

classification. The designed representation with a bag of visual words is utilized by the computer vision community in visual recognition issues. Image sampling representation with huge regions lessens the execution time of the image classification process. The designed approach was employed for categorizing the four heart view planes. The designed approach was efficient for addressing target issues in real-time setups. But, the error rate was not minimized through mid-level image representation.

A designed system was introduced by G. N. Balaji et al., (2016) to identify and diagnose dilated cardiomyopathy (DCM). The designed system was carried out for denoising and development. Left ventricular segmentation was performed in individual frames. However, the disease diagnosing accuracy was not improved by the designed system.

Table 1. Optimization methods based disease diagnosis and their limitations

Author and Year	Technique	Demerits/Limitations
Prabhpreet Kaur., et al., (2019)	Hybridized Neuro-Fuzzy genetic algorithm	The computational cost involved in disease diagnosis was not minimized by neuro-fuzzy based on genetic algorithms.
Esmat Rashedi., et al., (2009)	Optimization algorithm	The designed algorithm experienced low searching speed and many algorithms have been designed to address existing problems.
Seyed Ali Mirjalili., et al., (2012)	PSO and GSA	The error rate was not minimized by PSO and GSA.
Sudipta Ghosh., et al., (2019)	Multi-Detector Computed Tomography (MDCT)	The optimal feature selection and extraction were not concentrated to minimize the pattern recognition time.
Otavio A. B. Penatti., et al., (2015)	Mid-level image representation	The error rate was not minimized through mid-level image representation.
G. N. Balaji., et al., (2016)	Dilated cardiomyopathy (DCM)	The disease diagnosing accuracy was not improved by the designed system.

2.2. Gravitational Search and Heuristic Search Methods for Disease Diagnosis

Gravitational Search Algorithm (GSA) stimulated through Newtonian gravity theory in physics. Many GSA alternatives were introduced with gravity-

related theories. Modified GSA was introduced with the objectives like continuous, binary, discrete, and multi-objective GSA.

FGSA was classified to resolve combinatorial optimization issues. To attain the tuning of PSS metrics, optimization issues were addressed by the FGSA technique. Though the designed algorithm was robust, complexity was higher. A hybrid heuristic technique was presented by Mansour Sheikhan et al., (2015). Through considering different speech signal features and principal feature vectors, a rich medium-size feature set was built. The Hybrid GSA-BGSA-neural technique was compared for optimization. However, the computational cost was not minimized by the hybrid heuristic method.

The evolutionary correlated gravitational search algorithm (ECGS) was introduced by J. Jayashree and S. Ananda Kumar (2019) for choosing the optimal number of features. The designed method examined every diabetic feature consistent with correlation. But, time consumption for optimal feature selection was not minimized by the ECGS algorithm.

Table 2. Gravitational search and heuristic search methods for disease diagnosis and their limitations

Author and Year	Technique	Demerits/Limitations
Ali Ghasemi., et al., (2013)	Fuzzy Gravitational Search Algorithm (FGSA)	Though the designed algorithm was robust, the complexity was not minimized.
Mansour Sheikhan., et al., (2015)	Hybrid heuristic method	The computational cost was not minimized by the hybrid heuristic method.
J. Jayashree and S. Ananda Kumar (2019)	Evolutionary correlated gravitational search algorithm (ECGS)	The time consumption for optimal feature selection was not minimized by the ECGS algorithm.

Yirui Wang et al., (2019) designed an HGSA based on population topology. Three experiments were conducted to enhance outcomes of GSA, heuristic algorithm, and PSO through function optimization. Danilo Pelusi et al., (2018) designed NFGSA to attain improved results. Designed algorithm include same computational complexity O(nd). ‘n’ indicates the number of agents and ‘d’ indicates search space dimension. An intelligent system was employed to change the GSA metric on a revised version of GSA. NFGSA enhanced optimization results of GSA and PSGSA without additional cost. Though the computational cost was minimized by NFGSA, the error rate remained unaddressed during the optimal search.

Researchers carried out their research on GSA to address the optimization issues in various fields.

2.3. Machine Learning Technique for Diagnosis Performance Enhancement

Machine learning (ML) methods allowed systems to study from experience. ML indicates the system's ability to obtain and incorporate knowledge. ML is a huge discipline with essential statistic learning theories of processes. Designed learning algorithms were employed to determine the hidden regularities in online data. The machine learning techniques categorize existing knowledge and attain new knowledge via smartly recording and reasoning of data points.

A 3DE machine learning (ML) approach was introduced by Valentina Volpato et al., (2019) for automated determination of left ventricular (LV) mass. The accuracy of the designed approach was improved based on cardiac magnetic resonance (CMR) reference using automated software with manual alterations. CMR analysis was carried out. ML approach reiterated measurements and quantified. However, the accuracy was not improved by the 3DE machine learning approach.

A machine learning method was introduced by P. Mohamed Shakeel et al., (2019) for the segmentation process and lessen time consumption. FFANN discovers segmented edges for describing identified texture boundaries with FFANN weights. But, time consumption for segmentation was not minimized by the machine learning method.

A fully automatic and fast ECG arrhythmia classifier with a brain-inspired machine learning technique was designed by Miquel Alfaras et al., (2019). The designed classifier included low-demanding feature processing with a single ECG lead. The designed approach was compatible with an online classification that aligned health-monitoring wireless devices. The ensemble combination developed the parallelism to instruct the classifier with remarkable speed. Though the designed approach was more efficient, PR accuracy was not improved by a fast ECG arrhythmia classifier.

A new real-time machine learning system was introduced by Xiaochen Tang et al., (2020) for Arrhythmia categorization. But, complexity was not minimized by the machine learning system. A neighborhood component analysis (NCA) based linear transformation was carried out by Md Saiful Islam et al., (2019) to enhance the AF detection robustness. The classification

models were instructed by transformed signals. AF detection was enhanced with arrhythmic rhythm types. Though the robustness was improved, linear transformation failed to enhance the accuracy performance.

Artificial intelligence (AI) technology was introduced by M Alsharqi et al., (2018) to present accurate, consistent, and automated interpretation for echocardiography. A machine learning technique was designed to enhance the diagnostic performance of echocardiography. An AI-assisted interpretation was carried out for an echocardiogram. It was feasible to provide a fast, accurate, and consistent echocardiogram assessment. The designed algorithm quantified different features like. The diagnosis was carried out to validate the automation process. But, the error rate was not minimized by AI technology.

A neural networks classification method was introduced by Eva Volna et al., (2014) for ECG prediction depending on ECG signals to identify normal cycles and arrhythmias for diagnosis. Two detection systems were generated with a neural network application. The designed method was employed to preprocess and classify into many classes. The outputs from classifiers perform the predictive character. The time series prediction was carried out through fuzzy transform with linguistic IF-THEN rules. However, the classification accuracy was not improved by the neural network classification method.

A new method was introduced by Zhen Chao et al., (2019) to enhance the interpolation method. The neural network includes an input layer, three hidden layers, and an output layer. A patch matching technique chooses input variables of a neural network. The output data was attained through a trained neural network. But, the computational complexity was not minimized.

Rongjun Ge et al., (2019) build a paired-views LV network (PV-LVNet) to determine the LV multi-type indices from the paired apical analysis. Depending on Res-circle Net, PV-LVNet detected LV and cropped LV region of interest with location module and image resampling. However, the PV-LVNet failed to achieve higher accuracy for clinical cardiac function estimation with lesser complexity.

In various machine learning methods, the complexity of the working process was not reduced. As the machine learning techniques increased the robustness, the linear transformation was not carried out to improve the classification performance. The time consumption for segmentation was not reduced through machine learning techniques. In addition, the accuracy was not enhanced through the 3DE machine learning approach.

Table 3. Machine learning techniques and their limitations

Author and Year	Technique	Demerits/Limitations
Valentina Volpato., et al., (2019)	3DE machine learning (ML) approach	The accuracy was not improved by the 3DE machine learning approach.
Xiaochen Tang., et al., (2020)	The real-time machine learning system	The complexity was not minimized by the machine learning system.
Md Saiful Islam., et al., (2019)	Neighborhood component analysis (NCA)	Though the robustness was improved, linear transformation failed to enhance the accuracy performance.
Rongjun Ge., et al., (2019)	Paired-views LV network (PV-LVNet)	PV- LVNet failed to achieve higher accuracy for clinical cardiac function estimation with lesser complexity.
Eva Volna., et al., (2014)	The neural networks classification method	The classification accuracy was not improved by the neural network classification method.
P. Mohamed Shakeel., et al., (2019)	Machine learning method	The time consumption for segmentation was not minimized by the machine learning method.

2.4. Deep Neural Learning Method for Disease Prediction

Deep learning was a developing tool for examining the image partially because of the complex multi-view format. A convolutional neural network was employed by Ali Madani et al., (2018) to categorize the standard views depending on labeled images and videos that collected the range of real-world clinical variation. But, the filtering technique was not used for accurate classification performance and minimized the noise rate.

A data-efficient deep learning classifier was introduced by Ali Madani et al., (2018) for the prediction process in cardiology. An image echocardiographic view classification was carried out by a pipeline supervised model for binary left ventricular hypertrophy categorization. A semi-supervised generative adversarial network model was introduced. However, the error rate was not minimized by a data-efficient deep learning classifier.

A deep neural network (DNN) was introduced by Sushravya Raghunath et al., (2020) to forecast the future clinical event from ECG voltage-time traces. A blinded survey of cardiologists was described with discriminating features of normal ECGs. The Cox proportional-hazard technique described the hazard ratio for two predicted groups. Though the accuracy was improved, the time consumption was not reduced by DNN.

A deep learning-based novel 2-D CNN approach was introduced by Elif Izci et al., (2019) for accurate classification. The architecture was analyzed on Electrocardiogram (ECG) signals. ECG signals were partitioned into heartbeats. Every heartbeat was changed into 2-D grayscale images for CNN structure. A deep learning framework was introduced by Ali Isin and Selen Ozdalili (2017) to perform ECG arrhythmia diagnosis by categorizing the patient's ECG. Deep CNN was employed as a feature extractor. Extracted features were inserted into the backpropagation neural network for classification.

A deep learning architecture was constructed by Artzai Picon et al., (2019OHCA patients attained the minutes after arrest onset. Data partitioned the patient-wise for performing the classifier design. The designed architecture was evaluated with 1D-CNN based on VF-detection features and support vector machine (SVM). However, the complexity level was not minimized by the designed deep learning architecture.

A semi and fully automatic algorithm were introduced by Olivier Bernard et al., (2016) for partitioning endocardium in RT3DE data to obtain relevant clinical indices. A standardized evaluation framework was introduced to estimate and partition the LV border. A database includes multivendor cardiac ultrasound recordings in various centers with equal reference measures. However, the time consumption was not minimized by a semi and fully automatic algorithm.

A new probabilistic brake pressure estimation method was introduced by Chen Lv et al., (2018) for electrified vehicles with ANNs and Levenberg-Marquardt backpropagation (LMBP). The high-level architecture was illustrated for brake pressure assessment. Standard backpropagation (BP) algorithm was employed for the training of a feed-forward neural network (FFNN). Depending on BP, the LMBP method was introduced. A new approach was introduced by Patricia Melin et al., (2012). MNN and fuzzy system structures were introduced by genetic algorithms. The designed methodology was employed for human recognition. The optimal MNN was designed with genetic algorithms. Fuzzy system optimization for response integration, the genetic algorithm varied membership functions and rules. However, accuracy was not improved by the designed approach.

A user recognition method was introduced by Min-Gu Kim et al., (2020) with the deep learning method depending on the ensemble network after ECG signal transformation into two-dimensional (2D) images. A pre-processing process was carried out to eliminate the noise or distortion. The designed method was projected on 2D image space and converted into image data. A

deep learning-based ensemble network improves degraded performance from overfitting in a single network. But, the recognition time was not minimized by the deep learning method.

Table 4. Deep neural learning methods and their limitations

Author and Year	Technique	Demerits/Limitations
Ali Madani., et al., (2018)	Convolutional neural network	The filtering technique was not used for accurate classification performance and minimized the noise rate.
Artzai Picon., et al., (2019)	Deep learning architecture	The complexity level was not minimized by the designed deep learning architecture.
Olivier Bernard., et al., (2016)	Semi and fully automatic algorithm	The time consumption was not minimized by a semi and fully automatic algorithm.
Chen Lv., et al., (2018)	Probabilistic estimation method	The classification time was not minimized by the probabilistic estimation method.
Patricia Melin., et al., (2012)	Genetic optimization approach	The accuracy was not improved by the designed approach.
Min-Gu Kim., et al., (2019)	User recognition method	The recognition time was not minimized by the deep learning method.
Ali Madani., et al., (2018)	Data-efficient deep learning classifier	The error rate was not minimized by a data-efficient deep learning classifier.
Sushravya Raghunath., et al., (2020)	Deep neural network (DNN)	Though the accuracy was improved, the time consumption was not reduced by DNN.
Running He., et al., (2019)	The automatic arrhythmias classification method	The computational cost was not minimized by the automatic DNN classification model.

2.5. Convolution Neural Network-Based Disease Diagnosis

Echocardiography employed ultrasound equipment by Amirata Ghorbani et al., (2020) to gather high temporal and spatial resolution images. Deep learning employed echocardiography to find the local cardiac structures, to establish phenotypes for avoiding cardiovascular risk. EchoNet identified pacemaker presence, improved left atrium, left ventricular hypertrophy, left ventricular end-systolic and diastolic volumes. Computation complexity was not reduced by the deep learning method.

An automatic method was introduced by Suyu Dong et al., (2018) for LV segmentation. However, the error rate was not minimized by the automatic FCN method. An encoder-decoder deep CNN technique was introduced by Sarah Leclerc et al., (2019) for assessing 2D echocardiographic images. The designed technique was used for segmenting the cardiac structures and determining the clinical indices. The cardiac acquisition was carried out for multi-structure ultrasound segmentation for echocardiographic assessment. But, the filtering process was not carried out to improve the disease diagnosis performance.

An AF detection method was introduced by Chaur-Heh Hsieh et al., (2020) with end-to-end 1D CNN to improve detection accuracy and lessen the complexity. An efficient 1D CNN through key components of the convolutional block was employed to improve detection accuracy and to attain optimal hyperparameters of CNN. However, the computational time was not minimized by the AF detection method.

An ECG arrhythmia classification method was introduced by Jingshan Huang et al., (2019) with deep CNN. But, classification accuracy was not improved by the ECG arrhythmia classification method.

Dataset was preprocessed and segmented. The research was supportive for assisting the doctor to understand the type of arrhythmia of the patient. However, the time consumption was not minimized by the CNN 2D model. A deep convolutional neural network (CNN) architecture was introduced by Tanvir Mahmud et al., (2020) depending on temporal convolution. A realistic augmentation technique was introduced to provide a reduction in class imbalance through a variational pattern of wavelet denoised ECG data. Pointwise-Temporal-Pointwise Convolution (PTP) unit was presented. The depthwise temporal convolution with changing kernel size was included. Though the classification accuracy was improved, the computational cost was not minimized by efficient deep CNN architecture.

A generic morphological arrhythmias classification technique was presented by Weifang Sun et al., (2019) for the detection of ECG heartbeat. Gray-level co-occurrence matrix (GLCM) was designed for features vector description. CNN approach was employed to perform classification. A designed technique was employed for morphological arrhythmias detection. GLCM description extracted the shape features vector for different arrhythmias from ECG with better performance. But, the error rate was not minimized by the generic morphological arrhythmias classification technique.

An 11-layer deep convolutional neural network technique was designed by Saroj Kumar Pandey and Rekh Ram Janghel (2019) for MIT-BIH

arrhythmia data classification consistent with ANSI–AAMI criterion. In the CNN model, an end-to-end classification method was introduced without denoising process. The number of classification get minimized and identified to segment the QRS complexes. However, the time consumption was not minimized by the designed deep convolutional neural network model.

Table 5. CNN based disease diagnosis with their limitation

Author and Year	Technique	Demerits/Limitations
Amirata Ghorbani., et al., (2020)	Echocardiography employed ultrasound technology	The computation complexity was not reduced by the deep learning method.
Suyu Dong., et al., (2018)	Automatic method	The error rate was not minimized by the automatic FCN method.
Sarah Leclerc., et al., (2019)	Encoder-decoder deep convolutional neural network technique	The filtering process was not carried out to improve the disease diagnosis performance.
Tanvir Mahmud., et al., (2020)	Efficient deep convolutional neural network (CNN) architecture	The computational cost was not minimized by efficient deep CNN architecture.
Weifang Sun., et al., (2019)	Generic morphological arrhythmias classification technique	The error rate was not minimized by the generic morphological arrhythmias classification technique.
Jingshan Huang., et al., (2019)	ECG arrhythmia classification method	The computational complexity was not minimized by the ECG arrhythmia classification method.
Chunli Wang., et al., (2019)	ECG arrhythmia classification method	Time consumption was not reduced by the ECG arrhythmia classification method.
Jia Li., et al., (2018)	CNN model	The error rate was not minimized by a two-dimensional information vector.

An electrocardiogram (ECG) arrhythmia classification method was introduced by Chunli Wang et al., (2019) for ECG records to find nine types through 1D densely connected CNN. It was a convolutional neural network (CNN) model with better results in pattern recognition. However, the time consumption was not reduced by the ECG arrhythmia classification method. The morphology and rhythm of heartbeats were joined through Jia Li et al., (2018). The CNN model was employed to find irregular heartbeats or arrhythmias. The error rate was not minimized by a two-dimensional information vector.

2.6. Different Classification Methods with ECG Information

The classification method is a supervised machine learning method that studies the input data provided and classifies the new observations. Binary classifiers function with two classes or possible outcomes. Multiclass classifiers function with several classes to which the data belongs. The classification method is the key issue of pattern recognition where output value gets allocated to the given input value. Every data is allocated to one class label.

An ECG classification system was introduced by Boussaa Mohamed et al., (2015) for image processing and artificial intelligence. The descriptor haar-like was designed to enhance computation performance of haar features and multilayer perceptron classifier. Training and testing of the system were carried out using a learning base with labeled data and unlabeled data. The designed system improved the robustness and speed depending on the merits of the haar-like descriptor and artificial neuron networks. However, the computational cost was not minimized by the ECG classification system.

An electrocardiogram (ECG) arrhythmia classification technique was designed by Tae Joon Jun et al., (2018) with deep two-dimensional CNN. It provides better results in pattern recognition. Each ECG beat was converted as a two-dimensional grayscale image for the CNN classifier. The CNN classifier optimization comprised different deep learning techniques. But, the classification accuracy was not improved by the designed ECG arrhythmia classification method.

Tahmida Tabassum and Monira Islam (2016) designed a support vector machine (SVM) for identifying cardiac diseases. Though accuracy was improved, the time consumption was not minimized by the SVM method. An automated quality-aware electrocardiogram (ECG) beat categorization method was introduced by Udit Satija et al., (2019) for efficient ECG arrhythmias diagnosis. The designed method comprised three stages. But, time complexity was not minimized by the ECG beat classification technique.

A new morphological feature method was introduced by Hui Yang and Zhiqiang Wei (2020) for arrhythmias classification. An event of ECG signals was identified. The parametric features of ECG morphology were extracted from chosen ECG regions. A new feature was employed for examining the QRS complex morphology changes. The feature vectors were used by many classifiers for automatic diagnosis. However, the classification accuracy was not improved by the morphological feature method.

A new technique was introduced by Wenliang Zhu et al., (2019) for exact cardiac arrhythmias recognition and classification. But, the error rate was not minimized by cardiac arrhythmias recognition and classification method. The new system was introduced by Juyoung Park and Kyungtae Kang (2015) to identify arrhythmia for heartbeat classification and clinical decisions. HeartSearcher detects patients with arrhythmias from heartbeat categorization. The designed system reviewed every patient's typical heartbeat pattern in a regular expression. The patient gets ranked consistent with pattern similarities. Though the time consumption was minimized, the complexity cost was not minimized by the designed system.

A new group sparse representation model was introduced by Yi Guo et al., (2017) to partition cardiac chambers in fetal echocardiograms. The designed model was introduced to build up the inherent echocardiograms structure by reducing the group reconstruction error, sparsity, and distinctive terms in a unified framework. A new approach was introduced by G. N. Balaji et al., (2015) for heart muscle damage recognition. A new system was introduced with image denoising, enhancement, and segmentation for obtaining heart wall boundaries.

U2S parent network was instructed with U-Net network to recover ultrasound texture. S2U parent network generated ultrasound data from the indiscriminate input. An encoder network was connected to transform ultrasound images to sketch images. A fully integrated ECG signal processor (ESP) was employed by Nourhan Bayasi et al., (2016) for ventricular arrhythmia prediction. The adaptive technique was introduced for detection and the description of P-QRS-T waves was utilized to extract fiducial points. The designed method was robust to variation in ECG signal with high sensitivity and precision.

An arrhythmia monitoring system was presented by Shuenn-Yuh Lee et al., (2018). The designed system includes three parts. The first part was high-resolution for executing bio-signal sensing circuits. Feature of the circuit was minimal complexity, greater resolution, and lesser power consumption. A systematic training beat selection process was introduced by Xiaolong Zhai and Chung Tin (2018) with representative beats to enhance classification. The classification system was calculated for supraventricular ectopic beats (SVEB) and VEB detection. The designed method enhanced sensitivity and positive predictive rate. CNN classifier with routine training beats selection procedure achieved better results. The classifier was a personalized one. The designed system presented a consistent and automatic tool for arrhythmia heartbeat identification without feature extraction.

An energy-efficient electrocardiogram processor was introduced by Zhijian Chen et al., (2017) for arrhythmia detection. A sparse matrix computing design was developed to lessen computation complexity. But, the PSNR performance was not improved by an energy-efficient electrocardiogram processor. A new life-threatening arrhythmias detection algorithm was introduced by Felipe Alonso-Atienza et al., (2014) with ECG parameters through SVM classifiers. 13 metrics were determined for temporal and spectral features of ECG signal. A filter-type feature selection (FS) process examines the relevance of computed metrics. But, the detection time was not minimized by the life-threatening arrhythmias detection algorithm.

An efficient approach was introduced by Abdullah Jafari Chashmi and Mehdi Chehel Amirani (2019) for ECG arrhythmia diagnosis. Heartbeat categories were categorized. The designed system was employed to classify arrhythmia classes with greater accuracy. Though the accuracy was improved, the time consumption was not reduced minimized by an efficient approach.

Fajr Ibrahim Alarsan and Mamoon Younes (2019) introduced an efficient approach to categorizing the ECG signals. Each heartbeat was a mixture of action impulse waveforms created through specialized cardiac heart tissues. Heartbeat categorization was carried out for addressing the existing issues. The features were inputs of the machine learning technique. Spark–Scala tools reduced machine-learning (ML) performance. Though the accuracy performance was not improved, filtering performance was not improved by an efficient approach.

The wavelet transform was used for filtering and QRS identification in ECGs. A new robust deep neural network was introduced by Glenn Van Steenkiste et al., (2020) for ECG beat classification. An equine electrocardiogram (ECG) semi-automatic analysis was carried out by Glenn Van Steenkiste., et al., (2020) for ECG analysis. The wavelet transform was employed for filtering and QRS identification in ECGs. A new robust deep neural network was introduced for ECG beat classification. But, the computational cost was not minimized by wavelet transform.

An abnormal 12-lead electrocardiogram (ECG) was introduced for improved risk prediction. The majority of studies were performed with other populations and prior cardiovascular diseases. Henri K. Terho et al., (2018) represented the link of sudden cardiac death (SCD). But, time consumption was not minimized by a 12-lead electrocardiogram (ECG).

Table 6. Classification methods with their limitations

Author and Year	Technique	Demerits/Limitations
Udit Satija., et al., (2019)	Automated quality-aware electrocardiogram (ECG) beat classification technique	The time complexity was not minimized by the ECG beat classification technique.
Tahmida Tabassum and Monira Islam (2016)	Support vector machine (SVM) method	Though the accuracy was improved, the time consumption was not minimized by the SVM method.
Weifang Sun., et al., (2019)	Generic morphological arrhythmias classification technique	The time consumption was not minimized by the morphological arrhythmias classification technique.
Zhijian Chen., et al., (2017)	Energy-efficient electrocardiogram processor	PSNR performance was not improved by an energy-efficient electrocardiogram processor.
Felipe Alonso-Atienza., et al., (2014)	The life-threatening arrhythmias detection algorithm	The detection time was not minimized by the life-threatening arrhythmias detection algorithm
Fajr Ibrahem Alarsan and Mamoon Younes., (2019)	Heartbeat classification approach	Though the accuracy performance was not improved, filtering performance was not improved by an efficient approach.
Glenn Van Steenkiste., et al., (2020)	equine electrocardiogram (eECG) semi-automatic analysis	The computational cost was not minimized by wavelet transform.
Henri K. Terho., et al., (2018)	Abnormal 12-lead electrocardiogram (ECG)	Time consumption was not minimized by a 12-lead electrocardiogram (ECG).

From the above table, many researchers worked on different classification methods for pattern recognition. The problems identified from the existing techniques lack the performance efficiency of the proposed technique. They are higher time consumption during disease diagnosis, lesser peak signal-to-noise ratio, higher error rate, and higher computational cost.

3. Proposed Methodology

The proposed research techniques are introduced to perform pattern recognition from echocardiogram images with higher accuracy and less time and computational complexity. The research work is performed within various phases as shown in below Figure 1.

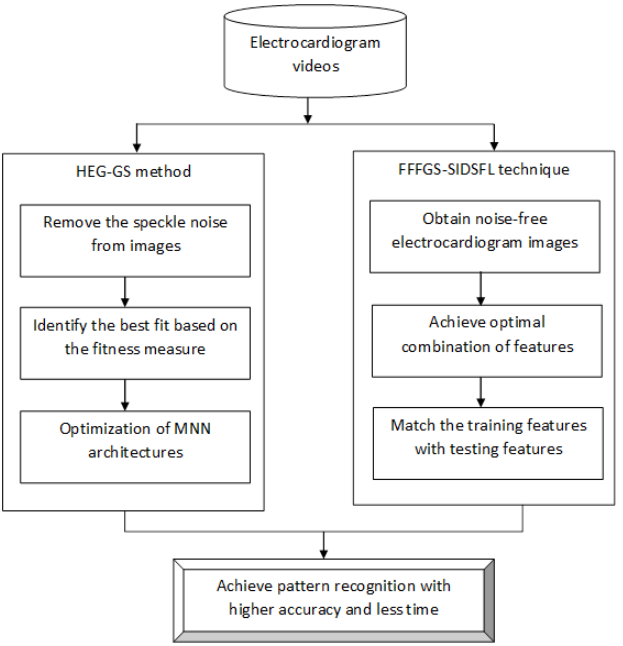


Figure 1. Flow process of the proposed methodology.

Figure 1 demonstrates the flow process of proposed techniques for performing efficient disease diagnosis through pattern recognition. A brief explanation of the proposed techniques is shown in the following sub-sections.

3.1. Hierarchical Elitism Gene Gravitational Search Method

3.1.1. Additive Kuan Speckle Noise Filtering Model

The first step in the design of the HEG-GS method is the Additive Kuan Speckle Noise Filtering model. By using this filtering model, the input images are preprocessed and thus increase the visual representation of images for disease diagnosis. The HEG-GS method takes echocardiogram videos with healthy and non-healthy patients as input. Then these input images are submitted to the preprocessing where the filtering process is carried out.

3.1.2. Hierarchical Elitism Gene GSO Optimization of MNN

After that, hierarchical elitism gene gravitational search is employed for pattern recognition with the assists of obtaining the best MNN. HEG-GS

method utilizes the MNN and architectures developed by the hierarchical genetic optimizer. Based on the divide and conquer principle, the modularity function is processed. By using MNNs, significant learning is achieved then the single Neural Network. In this algorithm, the number of preprocessed images is populated initially. Then the force of the agent is computed depending on the fitness measure. From that, the best fitness of the agent is detected to maximize the pattern recognition accuracy. Besides, the hierarchical elitism gene is employed in the gravitational search algorithm (GSA) for detecting the MNN.

3.2. Frost Filtration Fuzzified Gravitational Search Based Shift-Invariant Deep Structure Feature Learning Technique

The proposed FFFGS-SIDSFL technique is designed for providing accurate pattern recognition. The shift-invariant deep structure learning concept in the FFFGS-SIDSFL technique is implemented with three different steps namely preprocessing, feature extraction, and feature matching. By using these three steps, the pattern recognition is improved with less time and higher accuracy and peak signal to noise ratio.

Algorithm 1. Frost filterative fuzzified gravitational search based shift invariant deep structure feature learning

Input: Electrocardiogram database, Electrocardiogram videos,
Output : Improve pattern recognition accuracy

Begin
Given the Electrocardiogram videos into the input unit
For each video
Divide the frames $f_1, f_2, f_3, \ldots f_n$
For each f_i
. Apply filtering technique D
. Remove the speckle noise
End for
End for
For each preprocessed f_i
.Find an optimal combination of the feature vector
.Initialize the population of the agents
.For each agent $'i'$
.Measure mass $M_i\ (t)$ and fitness $fit_i(t)$
.Apply for fuzzy membership

.Find the best fit and worst fit

.Select best fit $\mathbf{arg}\,max\left(fit_j(t)\right)$

.Update the velocity $V_i^d(t+1)$ and position $P_i^d(t+1)$ of agent 'i' in 'd' dimension at time 't'

.Repeat the process until convergence is met

.End for

.Return (optimal combination of features)

.End for

.Extract the features, T, I, S, and C to create a feature vector 'F_v'

.If ($F_v = F_t$**) then**

δ_F returns '1'

.Learned features vector matched with the testing feature vector

Else

δ_F returns '0'

Learned features vector not matched with the testing feature vector

End if

Calculate the error R

Update the weights $\Delta\omega_1$, $\Delta\omega_2$, $\Delta\omega_3$

The process is repeated until finding minimum error

End

To start with the FFFGS-SIDSFL technique, several echocardiogram videos is utilized as input in the input unit. Then the input is partitioned into frames in hidden unit one. By taking these partitioned frames, an enhanced frost filtering technique is applied to eradicate the noise in the image at the hidden unit two. During this process, the noisy pixels are replaced by a neighborhood of all the pixels and thereby increase the peak signal to noise ratio. After that, stochastic gradient learning Fuzzified Gravitational Search (SGLFGS) is employed in the FFFGS-SIDSFL technique for feature selection and extraction. The best combination of features from the input is picked using SGLFGS. During this, the fitness is evaluated depending on how the feature combination is suitable to execute pattern recognition. The fuzzy triangular membership function is applied to determine the best fit and worst fit in SGLFGS. From that, the best fit is identified for each iteration from the population to increase the accuracy of pattern recognition. Subsequently, the velocity and position of each agent are updated to obtain the optimized combinations of features. From that, the texture, intensity, shape, and color features are extracted to create the feature vector. Then these feature vectors are given to the output unit where the Gaussian activation function is applied to match the feature vector. Here, a generated feature vector is validated with

a testing feature vector for detecting the occurrence of the disease from the electrocardiogram image.

The above algorithm shows pattern recognition with higher accuracy and minimum time. First, preprocessing in FFFGS-SIDSFL is employed to remove the speckle noise. The fuzzy implemented gravitational search algorithm is applied to select and extract the features from echocardiogram images. This process minimizes the pattern recognition time. In the last step, recognition is established by considering extracted features with maximum accuracy. As a result, FFFGS-SIDSFL carries out the disease diagnosis through pattern recognition with a reasonable time.

4. Simulations and Performance Metric Analysis

The experimental evaluation of two proposed techniques such as HEG-GS method and the FFFGS-SIDSFL technique is implemented in MATLAB. The proposed method utilizes echocardiogram videos taken from the https://e-echocardiography.com/page/page.php?UID=1429484681. The number of images is considered for conducting experiments is 15 to 150 for 10 iterations. The performance of proposed techniques is compared with existing PV-LVNet [1] and Neuro-Fuzzy based on a Genetic algorithm [2]. The parameters such as pattern recognition accuracy and computational time are considered to show the performance of the proposed techniques.

4.1. Measure of Pattern Recognition Accuracy

Pattern recognition accuracy is defined as the ratio of correctly recognized patterns (i.e., disease presence or absence) from echocardiogram images to the total number of echocardiogram images. It is measured in terms of percentage (%).

The performance comparison of the two proposed techniques is made with the existing PV-LVNet [1] and Neuro-Fuzzy based on a Genetic algorithm [2]. As compared to the other proposed and existing methods, the pattern recognition accuracy is effectively increased by the FFFGS-SIDSFL technique. This improvement is obtained through the shift-invariant deep structure learning in FFFGS-SIDSFL. The activation function is used for validating the learned feature vector with the testing feature. Therefore, the proposed HEG-GS method increases the pattern recognition accuracy by 2%

and 7% as compared to the existing [1] and [2] respectively. Besides, the proposed FFFGS-SIDSFL technique enhances the pattern recognition accuracy by 11% and 16% as compared to the existing [1] and [2] respectively.

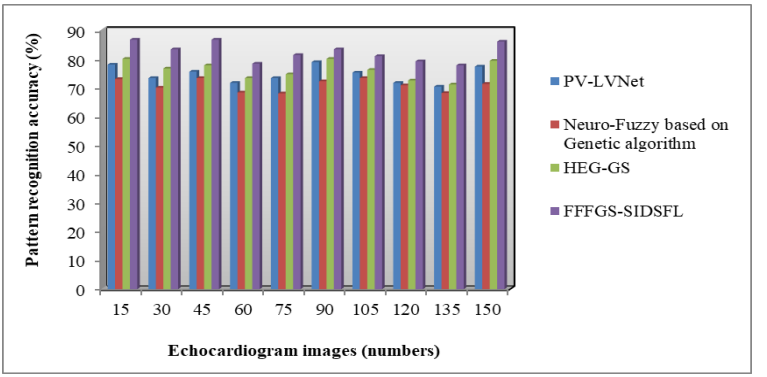


Figure 2. Performance of pattern recognition accuracy.

4.2. Measure of Computational Time

Computational time is referred to as the amount of time taken for pattern recognition from echocardiogram images. It is measured in the unit of milliseconds (ms).

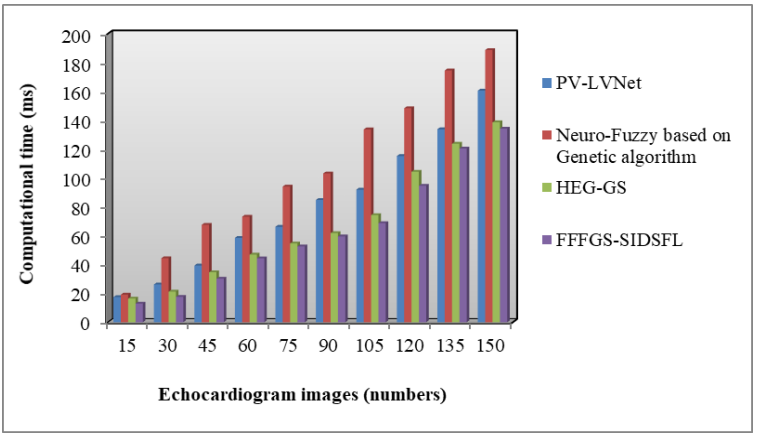


Figure 3. Performance of computational time.

Figure 3 demonstrates the experimental analysis of computational time based on different ECGs. The comparison of the proposed HEG-GS method and FFFGS-SIDSFL technique is made with existing [1] and [2]. From the analysis, the proposed FFFGS-SIDSFL technique results with minimum computational time using GLFGS where the optimal combination of features is only considered for pattern recognition. Therefore, the HEG-GS method reduces the computational time by 15% and 36% as compared to the existing [1] and [2] respectively. In addition, the FFFGS-SIDSFL technique minimizes the computational time by 23% and 42% as compared to the existing [1] and [2] respectively.

Conclusion

In our research work, three proposed techniques such as the HEG-GS method and FFFGS-SIDSFL technique are designed to achieve effective results in pattern recognition from echocardiogram images. At first, the HEG-GS method is introduced for improving pattern recognition using noise removing and MNN optimization. By using HEG-GS, the computational time is effectively reduced. Then the FFFGS-SIDSFL technique is developed to enhance the pattern recognition for early disease diagnosis. In the FFFGS-SIDSFL technique, the feature vector is generated using SGLFGS for detecting the occurrence of diseases. Through the FFFGS-SIDSFL, pattern recognition is performed with minimal complexity and higher accuracy. The simulation result shows that the proposed technique obtains higher pattern recognition accuracy with lower computational time.

References

[1] Rongjun Ge, Guanyu Yang, Yang Chen, Limin Luo, Cheng Feng, Heye Zhang, Shuo Li, PV-LVNet: Direct left ventricle multitype indices estimation from 2D echocardiograms of paired apical views with deep neural networks," *Medical Image Analysis*, Elsevier, Volume 58, 2019 Pages 1-12.

[2] Prabhpreet Kaur, Gurvinder Singh, Parminder Kaur, "An intelligent validation system for diagnostic and prognosis of ultrasound fetal growth analysis using Neuro-Fuzzy based on Genetic Algorithm," *Egyptian Informatics Journal*, Elsevier, Volume 20, Issue 1, March 2019, Pages 55-87 2018, Pages 1-33.

[3] Beatriz González, Fevrier Valdez, Patricia Melin, German Prado-Arechiga, "Fuzzy logic in the for the optimization of modular neural networks in pattern recognition," *Expert Systems with Applications*, Elsevier, Volume 42, 2015, Pages 5839–5847.

[4] Yirui Wang, Yang Yu, Shangce Gao, Haiyu Pan, Gang Yang, "A hierarchical gravitational search algorithm with an effective gravitational constant," *Swarm and Evolutionary Computation Base Data*, Elsevier, Feb 2019, Pages 118-139.

[5] Martha Pulido, Patricia Melin, German Prado-Arechiga, "Blood Pressure Classification Using the Method of the Modular Neural Networks," *International Journal of Hypertension*, Hindawi, Jan 2019, Pages 1-13.

[6] Sudipta Ghosh, Bhabani Prasad Chattopadhyay, Ram Mohan Roy, Jayanta Mukherjee, Manjunatha Mahadevappa, "Estimation of echocardiogram parameters with the aid of impedance cardiography and artificial neural networks," *Artificial Intelligence in Medicine*, Elsevier, Volume 96, 2019, Pages 45-58.

[7] Valentina Volpato, Victor Mor-Avi, Akhil Narang, David Prater, Alexandra Gonçalves, Gloria Tamborini, Laura Fusini, Mauro Pepi, Amit R. Patel, Roberto M. Lang, "Automated, machine learning-based, 3D echocardiographic quantification of left ventricular mass," *Echocardiography*, Wiley, Volume 36, Issue 2, 2019, Pages 312-319.

[8] Ali Madani, Ramy Arnaout, Mohammad Mofrad & Rima Arnaout, "Fast and accurate view classification of echocardiograms using deep learning," *Digital Medicine*, Volume 1, 2018, Pages 1-8.

[9] Long Teng, ZhongLiang Fu, Qian Ma, Yu Yao, Bing Zhang, Kai Zhu, and Ping Li, "Interactive Echocardiography Translation Using Few-Shot GAN Transfer Learning," *Computational and Mathematical Methods in Medicine*, Hindawi, Volume 2020, March 2020, Pages 1-9.

[10] Amirata Ghorbani, David Ouyang, Abubakar Abid, Bryan He, Jonathan H. Chen, Robert A. Harrington, David H. Liang, Euan A. Ashley & James Y. Zou, "Deep learning interpretation of echocardiograms," *Digital Medicine*, Volume 3, 2020, Pages 1-10.

About the Editor

Dr. R. Anandan

Dr. R. Anandan completed his UG, PG and Doctorate in Computer Science and Engineering.

He is IBMS/390 Mainframe professional and he is recognized as Chartered Engineer from Institution of Engineers in India and received fellowship from Bose Science Society, India. He completed 7 certification courses mainly from CISCO.

He is currently working as Professor and Head, Department of Computer Science and Engineering, School of Engineering and Associate Director – Innovation and Incubation, Vels Institute of Science, Technology & Advanced Studies (VISTAS), Chennai, Tamil Nadu, India which is Pioneer Institution in Engineering. He has vast experience in corporate and all levels of academics in Computer Science and Engineering. His interests include Artificial Intelligence, Soft Computing, Machine Learning, High Performance Computing, Big data Analytics, Image Processing, 3D printing and Knowledge Engineering. Under his eminent guidance, 8 PhD Research Scholars are pursuing degrees and 3 Research Scholars have been awarded PhDs.

He is a member of many reputed international and national societies like International Association of Engineers (IAENG) Hong Kong, Computer

Science Teacher Association (CSTA) New York, Universal Association of Computer and Electronics Engineers, USA, Association of Computer Sciences and Information Technology (IACSIT), Singapore, Institution of Engineers, India, Computer Society of India and Indian Society of Technical Education (I.S.T.E).

He serves as Editorial Board Member/Technical Committee/Reviewer of International Journals such Arab/UK/USA journals (Springer, Thomson Reuters, SCI, Elsevier).

He has published more than 130 research papers in various International Journals such as Scopus and SCI. He has presented 76 papers at various International Conferences. He received best reviewer award for the year 2017-2018 from The International Arab Journal of Information Technology (SCI Journal) under Elsevier Publication, Netherland and the reviewer of the year award for 2018–2019 from Energy Review Elsevier Publication, Amsterdam, Netherland. He received Appreciation award on Teachers Day for the academic year 2017-2018. He is the Recipient of the awards Dr. APJ ABDUL KALAM Best Scientist Award for Computer Science Engineering from BOSE Science Society, Tamil Nadu, India, Distinguished Researcher in Artificial Intelligence Award by Research under Literal Access Awards, India, Outstanding Scientist award by International Research Awards on Engineering, Science and Management, Award for excellence in Young Researcher in Computer Science Engineering by Global Outreach Research, New Delhi, India and Distinguished Professor Award by Vedant Academics Bangkok Awards – 2019 during International Award Conference on Multi-Disciplinary Research and Application - 2019 held in Kasetsart University, Chatuchak, Bangkok, Thailand. He received the Achievers Icon Award 2020 in the category Innovative Educator in Computer Science and Engineering by Band Opus India Pvt. Ltd. He authored and edited 17 books. Some of his leading books are *Software Engineering, Advanced Java Programming, Functions of Hardware and Internet, Bigdata and Hadoop Beginners Guide for OOPS Concepts in Java Programming, Web Technology, Statistics with R Programming, Information Security, Artificial Intelligence and A Closer look at Bigdata Analytics.* He filed 9 patents of his work. He is also associated as an Editor at Wiley, World Scientific Press and Nova Publishers.

Index

A

ad hoc network, ix, 207, 228
algorithm(s), v, viii, ix, 4, 9, 14, 15, 20, 24, 28, 36, 41, 42, 43, 44, 49, 51, 55, 56, 57, 60, 61, 63, 64, 70, 71, 72, 74, 75, 76, 78, 79, 80, 81, 83, 84, 86, 87, 88, 91, 95, 100, 101, 109, 112, 113, 123, 133, 137, 138, 141, 142, 143, 144, 146, 149, 156, 162, 163, 164, 165, 166, 167, 168, 169, 170, 173, 174, 175, 178, 179, 182, 183, 190, 195, 196, 198, 199, 200, 203, 206, 207, 213, 214, 215, 216, 217, 218, 227, 230, 231, 234, 235, 236, 238, 239, 240, 241, 243, 244, 249, 250, 252, 254, 256, 257
AODV, 208, 211, 219, 220, 227, 230
artificial intelligence, vii, viii, 1, 44, 59, 174, 247
artificial neural network, 93, 113, 138, 170, 172, 180, 192, 195, 198, 199, 204, 236, 257
authentication, viii, 19, 22, 24, 25, 26, 27, 37, 38, 43
authenticity, 21, 24, 25, 26, 33
automation, 46, 118, 241

B

bandwidth, 24, 26, 93, 210, 212, 213, 221, 230
bioinformatics, 84, 86, 88, 112, 113, 170
bloch sphere, 64, 65, 66
block chain, 19, 20
brix, 172, 191, 195
building blocks, 3, 56, 89

C

chromosome, 115, 116, 117, 118, 119, 120, 121, 123, 125, 126, 137, 138
chromosome classification, 115, 116, 117, 118, 119, 122, 138
classification, viii, 26, 83, 86, 112, 115, 116, 117, 118, 119, 121, 124, 125, 126, 137, 138, 173, 174, 181, 182, 190, 194, 196, 199, 200, 201, 203, 205, 206, 238, 240, 241, 242, 243, 244, 245, 246, 247, 248, 249, 250, 257
clustering, viii, 38, 84, 129, 141, 142, 144, 146, 160, 164, 165, 169, 181
color, 171, 172, 174, 176, 178, 179, 180, 181, 182, 186, 190, 192, 193, 197, 198, 201, 253
communication overhead, ix, 207, 210, 211, 212, 225
computation, 3, 9, 14, 15, 16, 19, 25, 27, 28, 80, 81, 109, 137, 212, 215, 233, 235, 246, 247, 249
convolution networks, 84
cryptography, 1, 15, 19, 20, 21, 22, 23, 24, 25, 27, 28, 29, 41, 63
cumulative distribution function, 225
cutin, 178, 189, 192, 193, 197
cybersecurity, 20, 30
cytogenetics, 116

D

data communication, 21, 208, 210, 215, 223, 224
database, 88, 98, 103, 108, 122, 125, 172, 176, 216, 229, 243, 252
deep learning, viii, 83, 84, 86, 88, 100, 101, 105, 107, 112, 175, 242, 243, 244, 246, 247, 257
deep neural networks, 84, 112, 256
deep structure feature learning, 234, 252
defects, 177, 178, 179, 190, 193, 194, 197, 198, 203
denoising, 118, 131, 132, 133, 137, 238, 246, 248
detection, 27, 30, 88, 127, 181, 182, 190, 198, 205, 206, 208, 209, 211, 212, 214, 228, 229, 233, 240, 241, 243, 245, 248, 249, 250
DNA, viii, 83, 86, 102, 103, 104, 111, 112, 127
drug discovery, 83, 84, 86, 91
dynamic clustering, 141, 146, 165

E

echocardiogram, ix, 233, 234, 235, 236, 237, 241, 250, 251, 253, 254, 255, 256, 257
electrocardiogram, 236, 246, 247, 249, 250, 254
encoding, 19, 21, 27, 30, 34, 36, 86, 107
encryption, viii, 1, 19, 20, 23, 24, 27, 28, 29, 34, 36, 64, 230
energy, 2, 5, 6, 38, 41, 45, 190, 226, 249, 250
engineering, vii, 5, 41, 54, 83, 86, 96, 101, 104, 115, 237
entanglement, 7, 8, 14, 15, 63, 64, 67, 69
extraction, 6, 86, 87, 88, 91, 100, 101, 102, 111, 116, 183, 193, 197, 234, 236, 237, 238, 248, 252, 253

F

feature selection, 237, 238, 239, 249, 253
fetal growth, 236, 256
food, 171, 175, 182, 205
formation, viii, 83, 84, 104, 105, 162, 163, 164, 165
fruits, 171, 172, 173, 174, 175, 176, 177, 179, 180, 181, 182, 200, 204
function estimation, 236, 241, 242
fuzzy membership, 252

G

gene expression, 127
grading, viii, 171, 172, 173, 174, 175, 176, 177, 178, 179, 180, 181, 182, 183, 185, 186, 190, 192, 193, 194, 196, 197, 198, 199, 200, 201, 202, 203, 204
graph, 102, 104
gravitational constant, 257
gravitational search, 234, 239, 251, 252, 254, 257
gravity, 7, 238
growth, 28, 41, 84, 88

H

histogram, 11, 12, 13, 181
hub, 93, 96, 97, 98
human, 30, 55, 89, 116, 138, 172, 231, 243
human body, 30, 89
hybrid, 239
hydrophobicity, 86, 89
hyperspectral analysis, 172
hypertrophy, 242, 244

I

image(s), 27, 28, 29, 50, 54, 87, 100, 115, 116, 117, 118, 119, 120, 121, 126, 131, 132, 133, 134, 135, 136, 137, 172, 173, 174, 175, 178, 180, 181, 182, 183, 184, 185, 189, 190, 192, 193, 194, 197, 198, 200, 202, 205, 233, 234, 235, 236, 237, 238, 241, 242, 243, 244, 245, 247, 248, 250, 251, 252, 253, 254, 255, 256
impulsive, 131, 132, 137

industry, vii, viii, 41, 42, 46, 51, 55, 57, 58, 204
information exchange, 6, 25, 46
information processing, 41, 112
information technology, 16, 21
input signal, 14
integration, 42, 44, 58, 243
intelligence, 59, 83, 84, 241
interface, viii, 23, 25, 31, 58, 63, 98, 106
interference, 6, 8, 16, 225
intermediaries, 222
issues, viii, 1, 14, 19, 22, 43, 44, 63, 95, 214, 224, 225, 228, 235, 237, 238, 239, 240, 249

K

k-means, viii, 141, 142, 143, 144, 146, 149, 161, 162, 163, 165, 166, 168, 169

L

Lagrange multipliers, 228
language processing, 112
learning, viii, 41, 44, 50, 52, 56, 58, 60, 83, 86, 87, 88, 100, 101, 102, 107, 109, 111, 112, 113, 174, 234, 236, 237, 240, 241, 242, 243, 244, 247, 249, 252, 253, 254, 257
light, viii, 8, 19, 20, 28, 36, 50, 65, 66, 67, 89, 178, 228
linear function, 78, 79
local thresholds, 137
logistics, 177

M

machine learning, 15, 41, 42, 43, 44, 45, 51, 52, 55, 56, 58, 60, 86, 112, 174, 175, 178, 182, 190, 203, 235, 240, 241, 242, 247, 249, 257
MANET(s), ix, 207, 208, 209, 210, 211, 213, 214, 226, 228
mapping, 66, 67, 95, 116, 228
mass, 86, 116, 119, 120, 237, 240, 252, 257
matrix, 27, 31, 67, 104, 109, 111, 129, 145, 153, 196, 199, 245, 249
maximum resemblance data labeling, 141, 144
media, 26, 29, 60, 214, 226
memory, 3, 5, 9, 14, 15, 29, 49, 93, 129
microarray technology, 127
mobile phone, 178, 183, 190, 192, 193, 198, 200
modules, 44, 45, 47, 98
molecules, 83, 84, 91, 102
morphology, 246, 247
motif, v, viii, 83, 84, 85, 86, 87, 88, 91, 100, 101, 102, 103, 104, 106, 107, 108, 109, 112, 113, 138, 170
multimedia, 208
multivariate analysis, 181
multivariate statistics, 205

N

near-infrared, 172, 173, 179, 181, 182, 191, 197, 198, 200
networking, 19, 21, 23, 25, 26
neural network(s), 59, 84, 88, 91, 93, 100, 101, 106, 107, 109, 110, 112, 113, 116, 175, 200, 235, 236, 237, 241, 242, 243, 244, 245, 246, 249, 256, 257
NIR spectra, 191
nodes, 20, 24, 34, 38, 110, 208, 210, 211, 213, 215, 216, 219, 220, 221, 224, 225
nodules, 222, 224
nucleotide sequence, 103
null, 149, 150, 151, 156, 158, 159, 184, 186, 189
null hypothesis, 184, 186, 189

P

parallelism, 80, 240
pattern recognition, 116, 118, 233, 234, 235, 236, 237, 238, 246, 247, 250, 251, 252, 253, 254, 255, 256, 257
pharmacovigilance data, 141
platform, 21, 22, 29, 50, 54, 93

PMS, v, 41, 42, 43, 44, 45, 47, 48, 49, 50, 51, 52, 53, 55, 56, 58
polar, 119, 208, 224
prediction, v, 27, 59, 60, 83, 84, 86, 87, 88, 91, 93, 100, 102, 106, 112, 113, 127, 128, 129, 130, 131, 138, 170, 181, 193, 195, 201, 202, 204, 205, 206, 241, 242, 248, 249
private information, 22, 29
programming, viii, 1, 8, 9, 42, 44, 45, 47, 91, 93, 94, 96, 130, 228
programming languages, 2
protection, 22, 23, 25, 26, 28, 30, 228
protein sequence, viii, 83, 86, 88, 93, 101, 107, 113, 129
protein structure, 84, 87, 88, 93
prototype, 9
PSO, 208, 228, 238, 239
Python, v, viii, 1, 8, 9, 10, 32, 94, 109

Q

Qiskit, v, viii, 2, 9, 10, 16, 63, 64, 65, 66, 67, 69, 75, 76, 77, 81
QKD, 19, 20, 25, 31, 36
QRS complex, 246, 247
quality of service (QoS), 207, 210, 213, 214, 226
quantification, 30, 257
quantum algorithm, viii, 1, 14, 15, 63, 64, 70, 72, 80, 81
quantum bits, 3, 4, 13, 19, 64
quantum computer, viii, 1, 3, 4, 5, 7, 8, 9, 13, 14, 15, 41, 44, 63, 64, 80
quantum computing, v, vii, viii, 1, 2, 3, 4, 6, 7, 9, 11, 13, 14, 15, 16, 19, 41, 42, 45, 49, 51, 57, 63, 64, 65, 80, 81
quantum cryptosystem, 19, 20
quantum mechanics, 1, 2, 3, 6, 7, 19, 20, 63, 67, 69
quantum state, 4, 5, 6, 7, 13, 14, 58, 64, 65, 67
quantum structure, 19
quantum theory, 25
qubits, 3, 5, 6, 7, 10, 13, 14, 15, 19, 20, 24, 31, 34, 64, 65, 68, 69, 74, 77, 79, 80

R

randon number generation, 208
recognition, viii, 116, 119, 122, 125, 126, 137, 138, 174, 186, 193, 194, 196, 202, 203, 204, 233, 234, 235, 236, 238, 243, 244, 248, 252, 253, 254, 256
reflectance spectra, 195
refractometer, 172, 191
regression, 191, 195, 208, 228
resources, 14, 15, 24, 210, 214, 221, 226
restoration, 22, 27
rhythm, 241, 246
rotation transformation, 115, 126
route creation, ix, 207, 208, 212, 213, 215, 216, 218, 220
routes, 207, 208, 212, 223, 227

S

scaling, 15, 115, 116, 118, 119
scanning tunnelling microscopy, 89
scattering, 91, 191
search space, 239
secure communication, 26, 41, 216
security, vii, 1, 4, 20, 21, 22, 25, 26, 28, 29, 30, 38, 45, 64, 93, 208, 210, 213, 214, 230
sensor(s), viii, 38, 39, 45, 171, 175, 178, 183, 191, 195, 198, 200, 201, 204, 206
signals, ix, 2, 4, 5, 21, 33, 115, 120, 241, 243, 247, 249
signal-to-noise ratio, 191, 235, 250
simulation, 11, 15, 81, 217, 218, 219, 220, 228, 256
software, 2, 3, 9, 14, 23, 32, 42, 45, 47, 53, 54, 56, 58, 93, 100, 146, 200, 240
solution, 28, 175, 176, 179, 183, 192, 196, 197, 199, 200, 204, 228, 237
specific gravity, 172, 204
spectroscopy, 179, 181, 182, 198, 205
standard deviation, 132, 186
stockpiling, 49, 96
storage, 29, 93, 127, 211, 212, 230
structure, viii, 23, 44, 83, 84, 86, 87, 88, 89, 90, 91, 96, 97, 100, 102, 104, 106,

112, 113, 118, 129, 180, 190, 209, 210, 234, 243, 245, 248, 252, 254
supply chain, viii, 172, 176
support vector machine, 112, 172, 243, 247, 250
sustainable quality, 172

T

topology, 23, 213, 215, 239
total soluble solids, 172, 178, 181, 191, 198, 202, 205
transactions, 24, 38, 45, 46, 48, 113, 170
transformation(s), 65, 73, 87, 100, 104, 115, 116, 118, 119, 120, 122, 190, 193, 240, 241, 242, 243
translation, 115, 116, 118, 119, 126, 137, 237
translation transformation, 118
transportation infrastructure, 21, 25
triangular membership, 234, 236, 253

U

ultrasound, 182, 233, 234, 237, 243, 244, 245, 246, 248, 256
uplink, 223
user data, 223

V

validation, 24, 26, 49, 236, 256
validation and privacy, 20
variables, 29, 34, 51, 57, 79, 195, 241
varieties, 129, 175, 183, 185, 186, 188, 192, 193, 198, 199, 201, 203, 204
vector, viii, 64, 65, 83, 86, 87, 91, 102, 106, 112, 118, 121, 123, 124, 129, 142, 172, 180, 208, 228, 229, 234, 236, 243, 245, 246, 247, 250, 252, 253, 254, 256

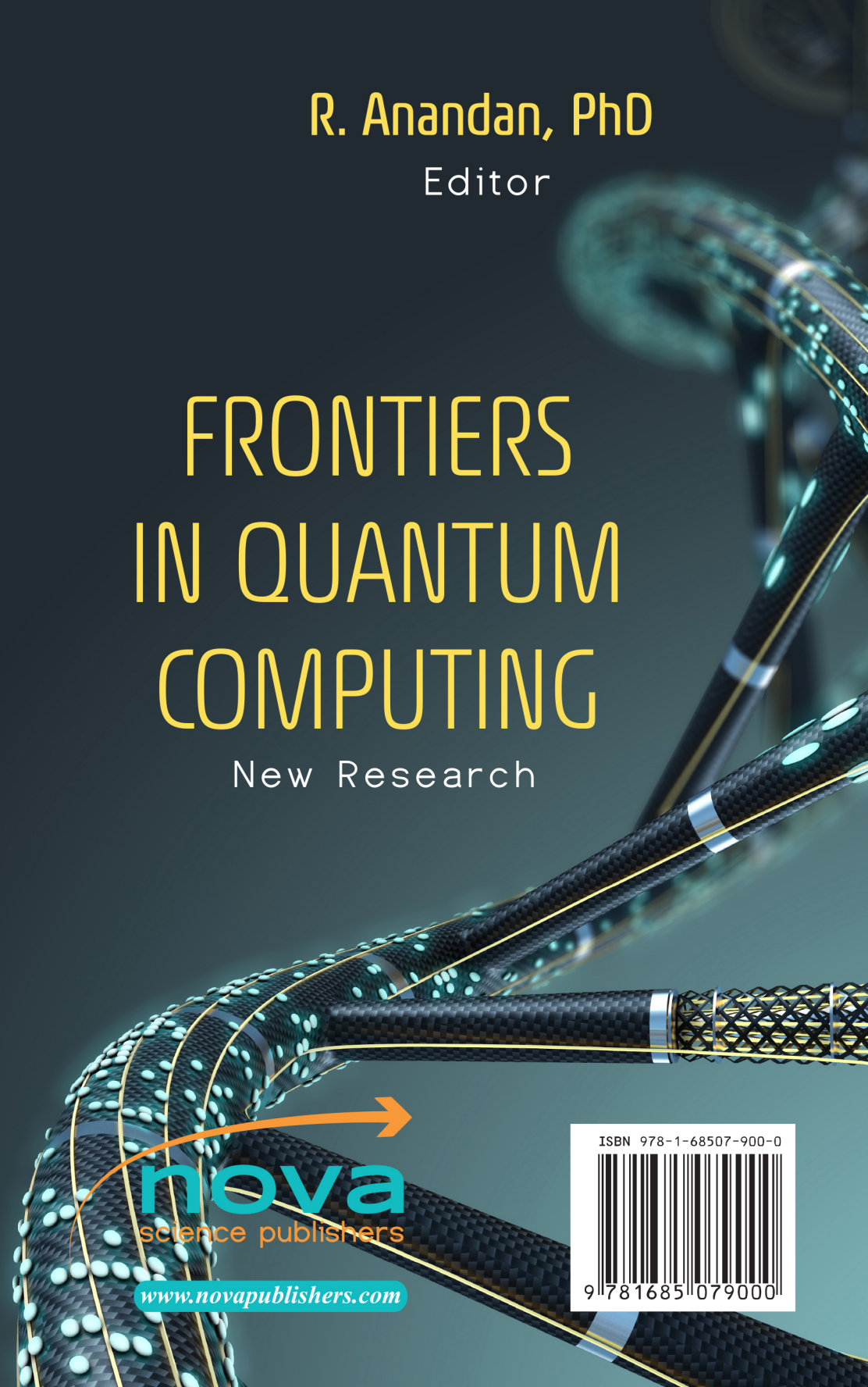
W

wavelet, 27, 115, 116, 118, 121, 125, 131, 132, 133, 137, 138, 245, 249, 250
wavelet transform, 115, 116, 121, 131, 249, 250
websites, 43, 50, 51, 54, 56
wireless devices, 228, 240
wireless networks, 22, 24, 38, 217
wireless sensor networks, 21, 38
wireless technology, 25

R. Anandan, PhD
Editor

FRONTIERS IN QUANTUM COMPUTING

New Research



nova
science publishers
www.novapublishers.com

ISBN 978-1-68507-900-0
9 781685 079000